LINGUISTIC ARCHAEOLOGY OF SOUTH ASIA

Franklin C. Southworth

RoutledgeCurzon
Taylor & Francis Group
LONDON AND NEW YORK

LINGUISTIC ARCHAEOLOGY OF SOUTH ASIA

Linguistic Archaeology of South Asia brings together linguistic and archaeological evidence of South Asian prehistory. Following introductory chapters describing the methodology of the interdisciplinary field known as "linguistic archaeology" and introducing the South Asian linguistic scene, the bulk of the book makes use of different types of linguistic evidence (such as loanwords, place names, regionally shared grammatical innovations) to draw inferences about prehistoric events. These inferences are then compared with and tested against the findings of other disciplines dealing with prehistory, including archaeology and palaeobotany. The final chapter uses the evidence of the previous chapters, in the light of the latest archaeological work, to sketch possible prehistories of the major South Asian language groups, and to suggest hypotheses for further investigation. The basic premise of the work is that ancient linguistic evidence, when put into its socio-historical context, can provide information to supplement the findings of other prehistoric disciplines. Thus the book is both a description of the unique methodology of "linguistic archaeology" and a treatment of South Asian historical linguistic data. It will be of interest to scholars and students of linguistics, archaeology, prehistory, and palaeobotany.

Franklin C. Southworth completed his PhD in Linguistics at Yale University. Subsequently, he taught Linguistics and South Asian Languages (Hindi, Urdu, Marathi, Nepali) in the South Asia Regional Studies Department at the University of Pennsylvania from which he retired in 1998. He spent over ten years in India doing fieldwork on Indo-Aryan (Marathi, Konkani, Hindi–Urdu) and Dravidian (Tamil, Malayalam) languages. His current research interest is SARVA (South Asian Residual Vocabulary Assemblage), an online dictionary of words of unknown origin in South Asian languages.

LINGUISTIC ARCHAEOLOGY OF SOUTH ASIA

Franklin C. Southworth

RoutledgeCurzon
Taylor & Francis Group
LONDON AND NEW YORK

First published 2005
by RoutledgeCurzon
2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

Simultaneously published in the USA and Canada
by RoutledgeCurzon
270 Madison Ave, New York, NY 10016

This edition published in the Taylor & Francis e-Library, 2005.

"To purchase your own copy of this or any of Taylor & Francis or Routledge's collection of thousands of eBooks please go to www.eBookstore.tandf.co.uk."

RoutledgeCurzon is an imprint of the Taylor & Francis Group

British Library Cataloguing in Publication Data
A catalogue record for this book is available
from the British Library

Library of Congress Cataloging in Publication Data
A catalog record for this book has been requested

ISBN 0-203-41291-5 Master e-book ISBN

ISBN 0–415–33323–7 (Print Edition)

CONTENTS

FIGURES

TABLES

PREFACE AND ACKNOWLEDGMENTS

This book is written for students of prehistory. While it inevitably contains technical linguistic details in places, I have done my best to make the work accessible to non-linguists. Chapter 1 is a discussion of the assumptions of the field I have dubbed "linguistic archaeology," and is intended for non-linguists. Each chapter concludes with a summary aimed at the general reader. Chapter 3 and Chapters 5–9 present various types of linguistic evidence which lead to specific prehistoric inferences about South Asia. Chapter 4 uses linguistic data from South Asia and other world areas to explore relationships between culture contact and linguistic change. Chapter 10 pulls together the various strands presented in the preceding chapters. The non-linguist interested in getting a quick view of the book is advised to start with Chapter 10, glancing at the sections cross-referenced there (including the chapter summaries) as needed.

If this work has an overall goal, it is to stimulate communication between historical linguists and other prehistorians, in the hope of finding ways to explore the totality of prehistoric cultures. This issue is taken up especially in Sections **1.7** and **10.3**.

I no longer remember when I started telling colleagues and friends that my book on linguistic archaeology "is almost finished." As with most books of its kind, the contents have been presented at numerous conferences and discussed many times during preparation, and suggestions and references have been incorporated from many sources. In addition, I have been shameless in asking for help when I needed it, and colleagues have been unhesitatingly generous in providing it. Thus it is more than possible that I have forgotten the names of some of these contributors, along with forgetting that their ideas were not originally mine. I hereby offer my sincere apologies to any whose ideas I have used without proper acknowledgment, and ask their forgiveness.

To those colleagues and friends who have, over the years, given their time to discuss my work with me, have looked over drafts of chapters, and have freely offered suggestions, criticism, and encouragement, I feel a debt of deep gratitude. I would particularly like to mention Mahadeo Apte, George Cardona, Madhav Deshpande, Peter Hook, Henry Hoenigswald, Ashok Kelkar, Bhadriraju Krishnamurti, Nick Masica, Joan Mencher, Asko Parpola, Edgar Polomé,

Gregory Possehl, David Stampe, Romila Thapar, Michael Witzel, Norman Zide, and Arlene Zide. I am also grateful to the following for advice, assistance, criticism, and permission to use published materials: Lloyd Anderson, H. S. Biligiri, George Erdosy, Walter Fairservis, Anne Feldhaus, Dorian Fuller, Peter Gardner, K. M. George, Hans Hock, Paul Hockings, Damir Kalogjera, Mark Kenoyer, Ken Kennedy, William Labov, Karl Lamberg-Karlovsky, David McAlpin, Richard Meadow, P. J. Mistry, Thomas Moylan, Prabodh Pandit, Jagdish Sharma, Prem Singh, Stanley Steever, K. V. Subbarao, V. I. Subramonian, Dell Hymes, and Vishnu-Mittre, among others.

The following organizations have graciously allowed me to quote materials from their publications without charge: Harvard University Press, M. S. University of Baroda, The Center for Applied Linguistics, Orient Longman, Thomson Publishing Services, Random House Group Ltd, Stanford University Press, The University of California Press, La Société d'Édition Les Belles Lettres, Cambridge University Press, Oxford University Press, Blackwell Publishers, The University of Chicago Press, The Linguistic Society of America, Faber & Faber, and Librairie d'Amérique et d'Orient Adrien Maisonneuve-Jean Maisonneuve Succ.

Finally, I thank those whose efforts made it possible for this work to finally see the light of day: Shubha Chaudhury, who worked on data collection; Maureen Fadem, who created the basic bibliography; Denise Miyahana, who helped me get organized when I needed it most; and Brandy Moore, who obtained the necessary permissions and patiently called my attention to many errors and omissions.

ABBREVIATIONS

Abbreviations for Sanskrit works in Chapter 3 and elsewhere are identical with those used in CDIAL (Turner 1966), and are not listed here. See the Bibliography for bibliographical abbreviations.

A	Ahom (Assamese)
AA	Austro-Asiatic
App.	Appendix
AV	Atharvaveda
B	Bangla (Bengali)
BCE	Before current era
Bi	Bihari
Br	Brahui
BRW	Black and Red Ware
Bur.	Burushaski
C	Concord suffix
CD	Central Dravidian
CE	Current era
Class.	Classical Sanskrit
CM	Central Munda
Dr	Dardic
Dt	District
E	English; eastern inscriptions of Ashoka
E–W	Eastern and southwestern inscriptions of Ashoka
F	French
G	Gujarati
Ga	Gadaba
Go	Gondi (Dravidian); Gorum (Munda)
Gu	Gutob
H	Hindi
H–U	Hindi–Urdu
IA	Indo-Aryan
IIr	Indo-Iranian
J	Juang

K	Kashmiri
Ka	Kannada
Kd	Konda
Kg	Kodagu
Kh	Kharia
KHA	Khandesh
Ki	Kui
Kl	Kolami
KM	Kherwarian Munda
Ko	Kota (Dravidian), Konkani (IA)
KOK	The Konkan
Kr	Koraga
Ku	Kudux (Dravidian); Kurku (Munda)
Kw	Kuwi
lex.	Sanskrit lexicographies
LP	Lahnda–Panjabi
M	Marathi; (**4.3** only) Malayalam; (**6.1** only) Midland Ashokan dialect
Ma	Malayalam
MAH	Maharashtra-Desh
MAR	Marathvada
MBh	Mahabharata
Md	Manda (Dravidian), Maldivian (IA)
mg	meaning
MIA	Middle Indo-Aryan
MK	Marathi–Konkani
Mt	Malto
MT	mother tongue
Mu	Mundari
N	Nepali; northwestern inscriptions of Ashoka
NBP(W)	Northern Black Polished (Ware)
ND	North Dravidian
NIA	New Indo-Aryan (modern Indo-Aryan)
Nk	Naiki
NM	North Munda
NP	Noun phrase
NW	Northwest; northwestern Indo-Aryan languages
O	Oriya
OB	Old Bangla (old Bengali)
OCP	Ochre Colored Pottery
OG	Old Gujarati
OI	Old Iranian
OIA	Old Indo-Aryan
OM	Old Marathi
OP	Old Persian
OT	Other tongue (i.e. not mother tongue)
P	Panjabi

Pa	Parji (Dravidian), Pali (MIA)
PAA	Proto-Austro-Asiatic
p.c.	Personal communication
PCD	Proto-Central Dravidian
PD	Proto-Dravidian
Pe	Pengo
PED	Proto-Elamo-Dravidian
PGW	Painted Gray Ware
PIE	Proto-Indo-European
PIIr	Proto-Indo-Iranian
Pkt	Prakrit
PM	Proto-Munda
PN	Pahari–Nepali
PND	Proto-North Dravidian
PPN	Proto-Polynesian
PSD	Proto-South Dravidian
PSD1	Proto-South Dravidian-1
PSD2	Proto-South Dravidian-2
R	Ramayana
R-	Reduplicated syllable (in Munda citations)
Re	Remo
RV	Rigveda
S	Sindhi; (**4.3** only) South Indian language
Sa	Santali
SD	South Dravidian
SD1	South Dravidian-1
SD2	South Dravidian-2
Si	Sinhala (Sinhalese)
Skt	Sanskrit
SM	South Munda
So	Sora
SOV	Subject–Object–Verb
SV	Samaveda
s.v.	*sub verbo* (under the word)
SW	Southwest; southwestern inscriptions of Ashoka
T, Ta	Tamil
Te	Telugu
To	Toda
Tu	Tulu
U	Urdu
V	Vedic
VAR	Varhad (Vidarbha, Berar)
VP	Verb phrase
YV	Yajurveda
W	Southwestern Ashokan inscriptions (Chapter 6)

FORMULAE AND SYMBOLS

$A \rightarrow B$	A becomes or changes to B
$A \leftarrow B$	A is (descended) from B
$A \rightarrow B/X___$	A changes to B when preceded by X
$A \rightarrow B/___X$	A changes to B when followed by X
$A \rightarrow \emptyset$	A changes to zero (is lost)
Σ	sentence

1

THE SCOPE OF LINGUISTIC ARCHAEOLOGY

Contents

1.1. Introduction

As archaeology studies ancient artifacts and attempts to reconstruct a socio-historical context for them, the field I have dubbed 'linguistic archaeology'[1] seeks to reconstruct a *sociolinguistic context* for ancient linguistic forms, that is, to make inferences from the language of ancient texts and/or linguistic reconstructions about the groups who spoke the language in question – including their geographical location, chronology, social and linguistic links with other groups, social stratification, material culture, and ideology. Thus linguistic archaeology can be considered complementary to archaeology itself, in terms of its goals. It also shares with archaeology a commitment to rigorous scholarship, in the sense that conclusions need to be verifiable by methods that are generally agreed upon.

The practitioner of linguistic archaeology draws on assumptions and evidence from five subfields of linguistics:

(1) historical-comparative linguistics (**1.2**)
(2) linguistic palaeontology (**1.3**)
(3) sociolinguistics (**1.4**)
(4) glottochronology/lexicostatistics (**1.5**)
(5) philology (**1.6**)[2]

The purpose of this chapter is to describe, for the scholar who is not primarily a linguist, the relevant assumptions of these subfields, and the types of evidence which can be derived from them. The process of correlating linguistic and archaeological evidence is discussed in **1.7**.

1.2. Historical-comparative linguistics

Historical linguistics is the *diachronic* (as opposed to *synchronic*) study of language, that is, the study of linguistic change, while historical-comparative linguistics is concerned with the prehistoric inferences that can be drawn from comparison of the forms shared by related languages. (The relevant terms are defined here.)

1.21. Assumptions

The important assumptions of this field which are relevant to the present discussion include the following:

(A) *Accountability*: All elements of a language can in principle be accounted for as inheritances, borrowings, or internal innovations (**1.21A**).
(B) *Change*: All linguistic elements are subject to specific types of changes (**1.21B**).
(C) *Regularity of sound change*: Sound change is assumed to act systematically within stateable limits (chronological, geographical, social, and linguistic).
(D) *Reconstructibility*: Shared inheritances in related languages allow the reconstruction of corresponding features of a single earlier language from which the related features can be said to be derived. *Related languages* are those

which show resemblances of a type which oblige the investigator to assume that both languages are descended from a single earlier language, that is, that they are historically altered forms of what was once the same language. (See further discussion in **1.22**.)

1.21A. Accountability

A central, though not often explicitly stated, assumption of historical linguistics is that the structural elements of a language (its sounds, lexical forms, and rules of word and sentence formation) can be, in principle, accounted for exhaustively either as (1) inheritances (continuations of previously existing elements), (2) borrowings (elements diffused from a different language or dialect),[3] (3) internal innovations, or as some combination of these three.[4]

(1) *Inherited forms*: Marathi **pāc** 'five' (phonetically [pa:t^s]) is inherited from Old Indo-Aryan (OIA) **pañca**, which is in turn inherited from a presumed Proto-Indo-European word with a form which may be represented as ***penk̂^{w}e** (cf. Latin **quīnque**, Greek **pénte**).[5]
(2) *Borrowed forms*: Hindi–Urdu **kitāb** 'book' is originally an Arabic word which entered Hindi–Urdu through the Persian of the Mughal invaders; Marathi **jankśan** 'road/rail junction' is borrowed from English.
(3) *Internal innovations*: (a) Marathi **-n** in **don** 'two' (probably on the analogy of **tin** 'three', since there is no inherited **-n** in the Old Indo-Aryan words for 'two' (OIA **dvau**, Hindi–Urdu **do**, etc.);[6] (b) Marathi **phaṭphaṭi** 'motorscooter, motorcycle'.
(4) *Inheritance combined with borrowing*: Marathi **āpan** 'you' (formal), 'self', 'we' (inclusive) is inherited from OIA **ātma(n)-** 'breath, soul, self'; however, the meaning 'we' (inclusive) is borrowed from Dravidian (see Southworth 1971; also Chapter 9, note 1).
(5) Borrowing combined with internal innovation: Marathi **jankśan** (see item (2)) has a second meaning in colloquial rural Marathi, namely 'sophisticated, urbane' (i.e. one whose sphere of knowledge and acquaintances goes beyond the village level), which appears to be an internal innovation, as in **jankśan māṇus** [lit. "junction man"] 'city slicker, sophisticated man, big shot'.

1.21B. Change

(*i*) *Types of linguistic change*. Inherited linguistic forms[7] are subject to various types of linguistic change, including those listed here.[8]

(1) *Sound change*: For example, consider the following:

   (a) Initial **tr-** of OIA (Old Indo-Aryan) changed to **t-** in most MIA (Middle Indo-Aryan) languages. For example, OIA **traya** 'three': Pali **tayo**, OIA **trika** 'triple'→ Pali **tiko** (this is part of a larger sound change; see **1.21C**).

(b) OIA **ś ṣ s** (palatal, retroflex, and dental sibilants) merged into **s** (dental) in most of MIA and NIA (see **6.12A** for exceptions in northwestern MIA and NIA; cf. also Bloch 1965: 68–74). For example, OIA **śata** '100': Pali **sata**, Hindi **sau**; OIA **satya** 'true': Pali **sacca**, Hindi **sac**; OIA **bhāṣā** 'speech': Pkt. **bhāsā**; OIA **bhāsa** 'light': Pkt. **bhāsa**; OIA **nāśa** 'destruction': Pali **nāsa**; OIA **nāsā** 'nose': Pali **nāsā**.

Sound change may result in the loss of a phonological contrast, either in a limited number of environments or all environments (case (b)), the disappearance of a phonological element (e.g. Latin **h** in the Romance languages, see **1.31**), the creation of new distinctions (usually secondarily),[9] or the movement of elements in phonological space without any change in the number of contrasting elements – such as the change of French /r/ from a tapped [R], still heard in southern French, to a uvular fricative [ʀ] (Chambers and Trudgill 1998), or the changes in the phonetics of vowels in Philadelphia (see **1.41C**).

(2) *Grammatical change*: This is defined as change in the underlying system of rules governing grammatical elements. Examples include the following:

(a) Changes in the order of grammatical elements, such as the changes in word order which occurred between Old English and modern English (compare the following Old English passage[10] and its modern English translation, especially the final clause: *Hēr on þisum ēare cōm Ānlāf and Sweȝen to Lundenbyriȝ on Nātīvitas Sancte Mārīe mid IIII and hundniȝontiȝum scipum, and hī þā on þā burh festlīce feohtende wǣron*… 'Here in this year came Olaf and Svein to London town on the Nativity of St. Mary with four and ninety ships, and they then on the town continuously attacking were…'.

(b) Loss of grammatical distinctions, such as the reduction of the declensional system of Latin which occurred in the Romance languages, reducing the forms of the noun from ten (nominative singular **stella** 'star', genitive singular **stellae**, dative plural **stellīs**, etc.) to two (singular **étoile** 'star', plural **étoiles**).

(3) *Semantic change* (i.e. changes involving the meanings which distinguish individual lexical items from each other): for example, OIA **bhakta**, the past participle of the verb **bhaj**, originally meant 'divided, distributed, received, enjoyed, consumed', and later came to mean 'cooked rice', which is the meaning of the NIA words derived from it (e.g. Bangla and Marathi **bhāt**). Three basic types of semantic change, all of which are especially important for the study of culture history, are given as follows:

(a) *Extension* of meaning, which involves the expansion of a word's meaning to include new referents which are seen as similar in form or function to the original referents: for example, the progressive extension of the

meaning of English **pen** (originally 'feather') to a quill pen, ink pen, fountain pen, ballpoint pen.

(b) *Narrowing* of meaning, the complementary process in which a word's meaning becomes restricted: for example, the restriction of the meaning of English **corn** (originally 'grain, seed', as in modern German and British English – cf. **acorn** 'oak-seed') to 'maize', probably by shortening of the phrase **Indian corn**; English **deer** originally referred to animals, especially quadrupeds (cf. German **Tier** 'animal', **Tiergarten** 'animal garden' = 'zoo'); English **meat** at one time meant 'food, nourishment'.

(c) *Transfer* of meaning: the English word **divan**, referring to a kind of sofa, comes from a Turkish word meaning 'council of state'; possibly the transfer took place by means of a word like Urdu **divānxānā** 'council room' which contained such furniture.

(*ii*) *Causes of linguistic change.* What causes linguistic change? The most general answer to this question is that we do not yet know enough to answer it. In order to begin to answer it, we need first of all to distinguish different types of changes. Some phonological changes, such as the merger of distinct phonological elements or the loss of elements (see examples in (i)), may be accounted for in terms of the principle of least effort. In American English, the distinction between the word-initial sounds written as **w (witch**, **wear**, **wine)** and **wh (which**, **where**, **whine)** has disappeared in a large area of the northern United States; the same is true of the distinction between **morning** and **mourning**, **born** and **borne**, **forth** and **fourth** (Labov 2001: 314–16). Such distinctions can apparently disappear without serious consequences since they have a relatively low *functional yield*, that is, the loss of contrast will not often lead to misunderstandings.

It should be noted, however, that this explanation fails to account for the fact that change may take place in one language and not in another, even when the conditions are similar. For example, the 'weakening' of voiceless stops like **p t k** to voiced stops **b d g** and subsequently to fricatives **v ð ɣ**, and in some cases to **∅** (zero), has occurred in a number of language families, including Romance languages (cf. Latin **recipio** 'take back': French **recevoir**, Spanish **recibir**), Indo-Aryan (OIA **kūpa** 'cave, well': Oriya **kūa**, Marathi **kuvā**), and Dravidian (as indicated by **p-v-∅** alternations like Tamil **tap-u, tav-ar̲u** 'die', **tā** (← ***tav-ā**) 'destruction'), leading to frequent cases of homonymy in a number of languages, whereas in many other cases intervocalic stops have survived unchanged for millennia (cf. Hawaiian **'ape** 'a plant' ← Proto-Polynesian **kape**, Hawaiian **kipa** 'to visit' ← PPN **tipa**).

Another type of change which might be attributed to the "least effort" principle is the simplification of morphological systems. This type of change we can see happening around us, in cases like the regularization of English irregular plurals (**oxes**, **louses**, **hoofs** instead of older **oxen**, **lice**, **hooves**) or irregular verb forms (**dreamed**,

dived, **wreaked** instead of older **dreamt**, **dove**, **wrought**). While the wholesale reduction of the conjugational and declensional systems of classical Latin which have occurred in the Romance languages, such as the loss of all the non-nominative case endings in nouns (see earlier), might be considered to lead to a simpler form of speech, they have also led to greater expenditure of energy in another respect: instead of Latin **stellae** 'of the star', a French speaker must produce the slightly longer phrase **de l'étoile**; instead of **vidēbāmur** 'we were seen', she must say **nous avons été vues** or **on nous a vues**. Again, we must ask what pressures kept the old morphological structures in place before they changed, and what keeps similar structures (e.g. the complex case system of Finnish, or the highly irregular verbal system of Pashto) in place today. Clearly the drive toward ease of physical effort or toward a simpler system of rules is not a sufficient answer.

Many semantic changes are related to changes in the real world, such as the creation or discovery of new items or activities which need to be named (e.g. **software**, **pacemaker**, **quasar**, **SARS** (severe acute respiratory syndrome)) and the obsolescence of older items: terms for sailing ships and horse-drawn vehicles come to mind. Changes in technology or socioeconomic circumstances may lead to extension, narrowing, or transfer of meaning: see the cases of **pen** and **corn** discussed earlier. Some changes may result from slippage in transmission between generations, or perhaps between different segments of the population, when words are used in contexts that allow more than one interpretation (see **deer**, **meat**, **divan** discussed earlier). Various types of changes occur through contact between different languages, or between different dialects of the same language; see Chapter 4 for an exploration of this subject.

Changes linked to language contact, apart from borrowed words, also include changes in phonological and grammatical structure, such as the change of numeral structures in Saurashtri from the Indo-Aryan type 24 = '4 + 20' to the Dravidian type 24 = '2 × 10 + 4' (see **4.26**) – note also older English **four and twenty**, which has the same pattern as the German **vierundzwanzig**, while the more modern English **twenty-four** is closer to the French **vingt-quatre**. Cases like Saurashtri provide evidence that changes of this kind can arise through language contact, since the grammatical structure of Saurashtri is almost identical to that of Tamil though its vocabulary is largely of Indo-Aryan origin; however, it is often difficult to prove that specific changes in the past resulted from language contact (see **3.3**).

It is generally accepted that discontinuities in communication between groups create the potential for the language of each group to "drift" in different directions. Perhaps the best example of this phenomenon is the Polynesian languages, consisting of different groups living on isolated islands, which had little or no contact with each other after their initial separation, until the recent availability of rapid means of travel and communication. The existence of both regional and social dialects is also evidence of this tendency – even though the separation in these cases is social more than geographical. It is probably reasonable to assume that any group of people who communicate with each other more than with other

people will share more linguistic features with each other than with members of other groups; that is, the linguistic usage of group members tends to converge. This seems to be true down to the level of nuclear families, married couples, and siblings,[11] and is also true of those professional groups who communicate frequently in contexts which exclude other members of the speech community, such as soldiers, academicians, musicians, police, prisoners, gamblers, gangsters, etc. This being so, the answers to questions about why languages begin to change at particular times, why these changes sometimes go on for generations, and why in other cases they stop, are most probably to be sought in the interactions between language and social life. (See **1.41C** for further discussion.)

1.21C. Regularity of sound change

Apart from certain sporadic changes, which usually affect individual words,[12] sound change is assumed to operate without exception within stateable limits of time, space, and linguistic conditions. For example, most word-initial consonant sequences[13] of OIA became single consonants in MIA sometime in the middle of the first millennium BCE, except certain sequences in the speech of the northwestern region. Thus

OIA **sthāna** 'abode' → Pali **thāna**
OIA **kṣaṇa** 'moment' → Pali **khaṇa**
OIA **svādu** 'tasty' → Pali **sādu**
OIA **dvīpa** 'island' → Pali **dīpa**.

The exceptions to the general statement can often be described in terms of specific geographical regions, time periods, linguistic environments, or social contexts.

Restriction by geographical region. OIA **tr-** and **dr-** were retained in much of the northwestern region during the MIA period, and have survived in some northwestern NIA languages. Thus

OIA **traya** 'three' (see earlier text) → **trayo** in the northwestern inscriptions of Ashoka (see Bloch 1950: 54); cf. modern Kashmiri **trih** and compare Hindi **tīn**;
OIA **dranga** 'frontier station' → NW MIA **drãga**, Kashmiri **drang**, but **ḍ-** in Marathi **ḍāngi** 'customs officer' (Turner, CDIAL 6615).

Restriction by time period. Though documentation is not available in many cases, linguists believe that sound changes are generally operative only during specifiable periods of time. Examples are as follows:

(a) in Latin, **s** between vowels → **r** (note **ae*r*-is**, **flō*r*-is**, **ope*r*-is**, genitives of **aes** 'copper', **flōs** 'flower', **opus** 'work') by the mid-fourth century BCE, a date established on the basis of textual study. After the completion of this change,

words which entered the language either as loans (e.g. **pausa** 'pause' ← Greek, **gaesum** 'spear' ← Celtic) or as new creations (e.g. **dēsuper** 'from above', **resēminō** 'beget, reproduce') did not undergo the change (Niedermann 1953: 94–5).

(b) In MIA the consonant combination **kṣ**, which had been frequent in OIA, merged in most cases with **k(k)h** or **c(c)h**, depending on the region. This change, whose beginnings are observable in Vedic Sanskrit, was completed before the middle of the third century BCE, since all words which had OIA **kṣ** appear in the Ashokan inscriptions with **k(k)h** or **c(c)h** (see **6.13B** and Bloch 1950: 56).

Restriction by linguistic environment

(a) The change of intervocalic **s** → **r** in Latin (see text before) did not take place in words which had **r** in an adjacent syllable (such as **rosa** 'rose', **miser** 'wretched').

(b) In a number of Indo-Aryan languages, long vowels of OIA and MIA were shortened in unaccented syllables, but retained their length in accented syllables (see examples and discussion in **5.22C**).

Restriction by social context

(a) Studies of linguistic change in progress have shown a number of examples of change limited to, or most prominent in, particular socially defined groups: for example, an early study by Labov of the speech of Martha's Vineyard (an island off the coast of Massachusetts) showed that some residents of the island had a tendency toward modifying the first element in the diphthongs /ai/ and /au/ (as in **light**, **house**), and that the strength of this tendency correlated with the individuals' identification with life on the island as opposed to the mainland (Labov 1972: 1–42); other examples and references can be found in this same source.

(b) Maxine Berntsen's study of the speech of Phaltan, a town in Maharashtra, shows several examples of linguistic differences between "standard" and "non-standard" speakers: for example, loss of the contrast between **n** and **ṇ** in non-standard speech; insertion of a vowel in initial consonant clusters in non-standard speech (e.g. **tarās** for standard **trās** 'trouble'); loss of aspiration after consonants in intervocalic position (e.g. **rādā** for **rādhā** 'Radha' [woman's name]).[14]

While many sound changes which have been studied have a residue of exceptions, the assumption of regular sound change is best retained as a heuristic principle, since many investigations of sound changes have repeatedly shown that careful examination of apparent exceptions can often lead to a more precise and accurate statement of the change (see **1.21D**; also Jankowsky 1990).

A related point is that a change might occur with complete regularity in one region or in one social group, and yet be distributed irregularly in the larger

society of which that region or group is a part. Historical linguists make a distinction between *sound change* (or *phonemic change*) and *lexical diffusion* of change. As William Labov has pointed out, there are many known cases – both contemporary and in the past – of phonetically gradual, regular changes which have gone to completion without exceptions, such as the change of **s** → **h** → **∅** (zero) in New World Spanish, and the devoicing (**d** → **t**, **v** → **f**, etc.) of word-final consonants in English (Labov 1981: 302). At the same time, there is ample evidence from languages the world over for the diffusion of sound changes on a lexical basis; this can be illustrated by words like **bog**, **cog**, **dog**, **fog**, **hog**, **log** in American English: in eastern Massachusetts, some speakers pronounce all of these words with the vowel of **hawk**, while in many parts of the midwest and west all are pronounced with the vowel of **hock**; in intermediate areas such as western New England, New York, and New Jersey, various distributions of these two vowels occur in this set of words. (See Labov 1981: 269–72 for further discussion and references.)

1.21D. Reconstructibility

Given two or more related languages (defined in **1.2(D))**, knowledge of the possible types of linguistic change makes it possible to reconstruct portions of the original language (or *proto-language*) which is believed to have formed the latest common stage in the history of the languages in question. For example, the following comparisons allow the reconstruction of initial ***k-** in Proto-South Dravidian, with the assumption of a change ***k-** → **c-** before **i** (notation: **k** → **c**/#__**i)**[15] in Tamil and certain other languages:[16]

Tamil	*Kannada*	*Proto-SD*	*DEDR No.*[17]
kaṇ 'eye'	**kaṇ**	***kaṇ**	1159
kāṭu 'forest'	**kāḍu**	***kāṭu**	1438
kuṇṭam 'pit'	**kuṇḍam**	***kuṇṭam**	1669
kol- 'kill'	**kol-**	***kol-**	2132
ciri- 'laugh'	**kiri-**	***kiri-**	1562
cilumpu 'verdigris'	**kilu(m)bu**	***kilumpu**	1586

To what extent can historical linguists claim to reconstruct reality? This question has been much debated, and there is perhaps no current consensus. At one extreme, reconstructed sounds may be regarded as merely a convenient shorthand to represent a set of sound correspondences, such as Tamil **c**: Kannada **k** in this example. For example, Proto-Dravidian ***ẓ** can be used to represent the set of corresponding sounds which occur in cognate sets like the following: Tamil **kīẓ** 'low, below', Kota **kīṛ**, Kodagu **kīḷ**, Parji **kiṛi**, Kudux **kiyyā**, Brahui **kī**, etc. (DEDR 1619; Emeneau 1970: 98ff.). At the other extreme, a reconstructed form may be given nearly the status accorded to attested items, especially when reconstructing

forms of an attested language: for example, Turner reconstructs an OIA form ***trāyaḥ** 'three', alongside of the attested **trayaḥ**, on the strength of an Avestan form with **ā** as well as modern forms like Tirahi **trā** (see Turner CDIAL 5994). In the present work, reconstructions are presented as the author's estimate of the most probable phonological shape of the proto-form.

There is one important reservation. As the above mentioned example shows, reconstruction depends on comparing items (e.g. words or sounds) in two or more different languages and reconstructing a single source for each item. Thus no systematic reconstruction is possible without the assumption of regular sound change. Workers in this field recognize that this *regularity principle* is an ideal notion, what Anthony Fox (1995: 140) has called a "necessary idealization." Thus we reconstruct uniform invariant proto-languages – without regional or social dialects, and without differences between formal and informal ways of speaking – even though we are aware that no such language has ever been seen in real life, and has probably never existed except in our reconstructions.[18] Presumably such reconstructions do no harm, as long as we recognize them for what they are, idealizations. (See further discussion in **1.31**.)

1.22. Types of inferences in historical-comparative linguistics

Resemblances in linguistic form can provide evidence of connections between different *speech communities* (defined below), existing in different places and at different times. Where such resemblances cannot be attributed to chance, they can show the existence of one of the following:

(1) *Continuity* of communication over time and space, as in the case of Old English → Middle English → modern English, or Old Indo-Aryan → Hindi, or French → Haitian Créole, implying a (linear) *genetic relationship* (**1.22A**).

(2) *Divergence* of communication through time and space (e.g. Latin → French, Italian, Spanish, Rumanian, etc.), implying (lateral) *genetic relationships and subrelationships*. (Where two or more laterally related languages are seen to be more closely related to each other than to a third, as in the case of French and Spanish *vis-á-vis* Rumanian, they are said to share a *subrelationship*; see **1.22B**.)

(3) *Links between* distinct speech communities: *non-genetic contact* (linguistic borrowing, linguistic diffusion, linguistic convergence; see **1.22C**).

These three principal types of prehistoric inferences which can be drawn in historical-comparative linguistics are discussed in more detail in **1.22A–C**.

Speech communities. For many purposes it may be sufficient to define the *speech community* as "a group of people who interact by means of language"

(Bloomfield 1933), or in a more sociolinguistically aware mode, "a group who share the same norms in regard to language" (Labov 1972:158). For the present purpose some additional specifications are necessary:

(1) Since we are dealing with prehistoric or early historic societies before the advent of mass communication, in which written language played at most a minor role, the linguistic norms shared by members of the community are norms of *spoken language*.
(2) We are dealing with *contiguous* groups, that is, groups of individuals who interact face-to-face on a regular basis – not, for example, entities like Bloomfield's "English speech community," which would include the entire English-speaking populations of North America, Britain, Australia, etc.[19]
(3) Like the terms *language* and *dialect*, the term *speech community* is necessarily vague in scope. If members of the same speech community share "the same norms" in regard to speech, how finely should these norms be defined? Labov, for example, speaks of older and younger speakers in New York City as belonging to 'slightly different speech communities' (1972: 158). Presumably, 'slightly different' may be taken to imply that with a looser definition of the shared norms, these two groups could be considered to belong to the same speech community. The relevance of this point here is that we need to recognize that this sort of indeterminacy is part of human social groups and human language. Thus when we attempt to reconstruct the language of a speech community, we do not know what part of the total original complex our reconstructions represent – or whether, perhaps more likely, they present a composite of different times and places. See further discussion in **1.31**.
(4) Regional dialects are not discrete entities (see **1.21C** end). In a large *dialect continuum* (defined in **1.22B**), such as the Hindi-speaking area of north India, or nineteenth-century Germany, communication between adjacent villages is usually without difficulty, but individuals living at greater distances might not be able to communicate easily. The same type of situation may have existed in prehistoric times, for example in the greater Indus Valley (see **10.21B**). Thus, assuming that every individual belongs to (at least) one speech community, it must be recognized that speech communities potentially overlap and do not have clear boundaries. Furthermore, individuals can belong to more than one speech community, even without bilingualism: for example, many illiterate South Asians speak both a village dialect and a variety of wider communication, that is, some form of a regional or sub-regional standard dialect. In prehistoric times, the dialects of market towns may have filled such wider communication roles. Here again, then, it is important to recognize that we are unlikely to be able to reconstruct more than a small portion of the linguistic complexity that actually existed.

1.22A. Genetic relationship

When we say that two contemporary languages are (genetically) related, we mean that both are historical continuations of a language spoken by a single earlier speech community.[20] In order to demonstrate such a relationship it is necessary to show **shared inheritances** between the languages, that is, formal correspondences which are not due either to chance or to diffusion (either from one language to the other, or from a third source). Satisfactory proof depends on both the quantity and quality of the etymologies. Examples of the kind of data used for proving relationships are given in **1.21D** (South Dravidian) and in **2.21A** (Indo-European). For further examples see **3.22A** and **8.23**.

The term *genetic relationship*, borrowed from biology in the early days of historical-comparative linguistics, is perhaps unfortunate in that it may suggest a possible connection between language and race (currently a controversial notion) or ethnicity. The term is used here, as it usually is in contemporary linguistics, merely to refer to common origin, as defined in the preceding paragraph. The same applies to semantically related terms like *mother language*, *parent language*, *daughter language*, *ancestor*, *descendant*, etc., which also embody the biological metaphor.

It is not possible to specify the *number* of comparable items needed to show that two languages are related. Brahui shares just over 300 lexical items with the rest of Dravidian (see the Brahui index in DEDR, pp. 756–8), and not even all of these are unquestioned cognates; yet the relationship of Brahui to Dravidian is not doubted by any serious scholar nowadays, because of the *nature* of the compared items. Though in this case the number of items is on the low side, the acceptance of Brahui as a member of the Dravidian family rests on the presence of words like personal pronouns, interrogatives, a few kin terms, verbs denoting basic concepts such as ‘be’, ‘burn’, ‘sleep’, and other basic non-cultural items such as ‘water’ and ‘milk’.[21] In fact, until Bray published his analysis of the language in 1909, Brahui was very poorly known, and was believed to be an Indo-Aryan language (Bray 1909: 18).

The vocabulary of Brahui, as known from the early twentieth century, is thus overwhelmingly of non-Dravidian origin (mostly from the neighboring Balochi and Sindhi languages), and the Dravidian element in Brahui is perhaps close to the minimum needed to prove a relationship. The proof thus depends on the assumption, shared by most historical linguists, that certain parts of a language are less likely to be replaced by diffusion or borrowing from another source – and that these parts include (1) words denoting universally human (or non-cultural) concepts, especially those which are high in frequency like personal pronouns and certain verbs, and (2) grammatical forms, such as a verb used in negation (Tamil **al**, Brahui **all**, DEDR 234).

To say that Brahui is related to Dravidian (or that “Brahui is a Dravidian language”) means that we consider it proven that there was at some time in the past a single speech community, which we can call Proto-Dravidian for lack of

a better term (see Chapter 8 for further discussion), and that the Brahui language of today represents a historical continuation of the speech of (some segment of) that community. This does *not* mean that all (or even any) of the present-day Brahui speakers are *necessarily* descended directly from individual speakers of proto-Dravidian (though it may often be the case that some significant proportion are so descended). Nor does it mean that the cultural or ethnic identity of the proto-speech community has been transmitted intact to the later community, since these questions must be investigated independently of the linguistic relationships. Thus it is inappropriate to use terms like "the Dravidians" – or for that matter, the "Aryans" (or the "Indo-Aryans") in connection with such proto-speech communities. The most neutral term would be "speakers of Proto-X," or the "Proto-X speech community."

This notion of a genetic relationship depends on an assumption which has not yet been stated, namely that the transmission of a language from one generation to the next proceeds in a fashion which leaves contemporaries able to communicate easily with one another. A naive form of historical-comparative linguistics would assume that transmission is exclusively from parents to children. What then of the so-called "mixed" languages, such as pidgins and creoles? While this is an enormous and complex subject which cannot be dealt with here, some points need to be mentioned. Haitian Creole may be taken as an example. When we say French → Haitian Créole, that is, only part of the story, since Créole contains other elements besides those which came to it from French. Without going into the controversy over the process by which such languages form, we can say that in such cases there has been a type of transmission which involved different speech communities – communities which perhaps had no shared language before the formation of the pidgin/creole. To say that Créole is a historical continuation of French is true to the extent that during its formation, some groups (which may themselves have lacked a shared language) were communicating in some fashion with speakers of French, that is, they were (partial) members of the French speech community. A full historical characterization of Créole would attempt to recognize all the other elements – lexical, grammatical, and phonological – which enter into its makeup. Such a description would give the history of the language in terms of the different sociolinguistic groups who contributed to its formation.

The example of Brahui is similar. Since the Dravidian element in Brahui is quantitatively overwhelmed by those parts of the language derived from Iranian and Indo-Aryan languages, it would be appropriate to say that Brahui is a historical continuation of Dravidian, Balochi, Sindhi, etc. And it is probable that the Brahui speech community has in fact absorbed many individuals from other groups over time (see Barth 1969). In this case, to say that Brahui is Dravidian is to recognize the Dravidian element as genetically the oldest segment of the language. (See Chapter 4 for discussion of other cases of this type in South Asia.)

1.22B. Genetic subrelationship

Historical-comparative linguistics recognizes two principle mechanisms by which a single uniform speech community may become differentiated, ultimately diverging into two (or more) communities with mutually unintelligible speech. In the simplest case, some members of a community migrate to a distance great enough to interrupt regular communication with the rest: the Malayo-Polynesian languages are classic examples; South Asian cases include Sinhala, the Indo-Aryan language of Srilanka, and the Gypsy languages. When such an event occurs, linguistic change proceeds independently within each segment of the original community and its descendants, leading ultimately to mutual unintelligibility. In such a case, linguists speak of the original speech community *branching* (or *splitting*, or *diverging*) and giving rise to two related languages. If such languages later undergo further splitting, it is often possible to reconstruct the original situation through the investigation of *exclusively shared innovations*. As an illustration, Figure 1.1 represents the history of three hypothetical related speech communities A, B, C which are descended from Proto-ABC. If A and B share one or more changes which are not shared by C, we infer that A and B are descended from an intermediate stage Proto-AB, in other words A and B belong to the same subgroup.

The linguistic statement that "A and B belong to the same subgroup" yields the historical inference that A and B are historical continuations of the language of a single speech community (the speakers of Proto-AB), which was intermediate in time between Proto-ABC and the present, and was at that time sociolinguistically different[22] from community C. Inferences of this kind make it possible to establish *linguistic stratigraphy*, as indicated schematically in Figure 1.1: at $time_0$ the unified Proto-ABC existed; at $time_1$, Proto-AB and C; at $time_2$, three separate speech communities have differentiated themselves from each other.

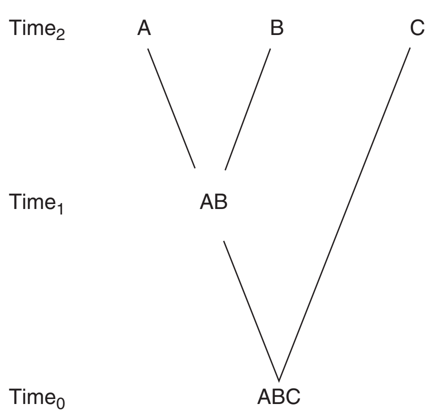


Figure 1.1 Illustration of subgrouping. (Languages A and B, but not C, belong to the same subgroup.)

However, differentiation often takes place without overt splitting, when a speech community becomes dialectally differentiated as a result of local innovations. Over a long period of time, the cumulative effects of such innovations can lead to a result similar to that found in the case of splitting, to the extent that individuals residing at a distance from each other no longer speak mutually intelligible varieties of the language – even though communication may be continuous between adjacent points (e.g. neighboring villages) throughout the region. In essence, this is what linguists believe happened to Vulgar Latin when it spread as a spoken language over Italy, Gaul, Iberia, etc., ultimately differentiating itself into what we now know as the Romance languages. The modern Indo-Aryan languages also developed by essentially the same process (but see Chapters 5 and 6 for a discussion of some of the intermediate stages). Such a situation, known to linguists as a *dialect continuum*, often continues to exist even after official state languages, usually based on the speech of urban elites, have been imposed over large areas. This was the normal state of affairs in many long-settled areas of Europe and Asia, up to the mid-twentieth-century CE.

When linguistic changes spread from one part of a speech community to another, those parts which share a number of changes can be said to belong to the same *dialect area*. Over long periods of time, however, successive linguistic changes tend to have (slightly or greatly) differing boundaries, reflecting historical changes in trade relations, shifting political allegiances, migrations, etc.[23] Thus it may happen that a single area shares changes with two or more adjacent areas, which themselves do not share changes with each other.

Figure 1.2 shows a simple case of this kind: the hypothetical dialect maps at the top of both diagrams depict a situation in which two different changes have taken place in adjacent areas, with an overlap in the intermediate region.[24] This overlap

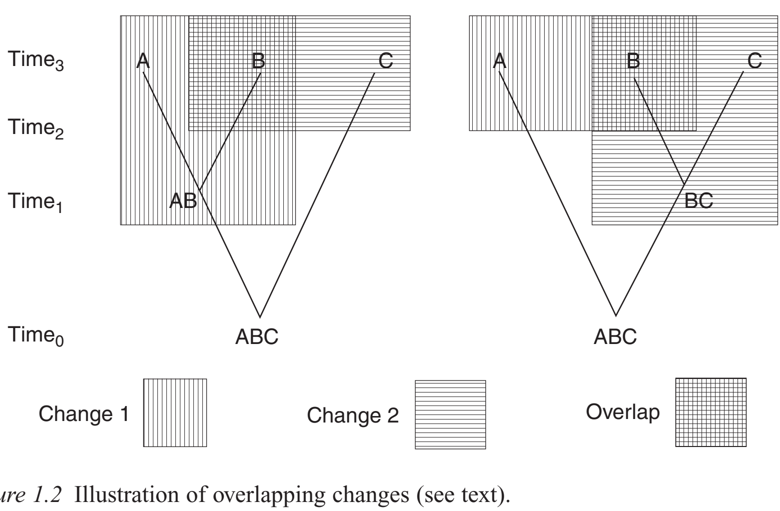


Figure 1.2 Illustration of overlapping changes (see text).

can be the outcome of two different historical processes: the lefthand diagram illustrates a case where change 1, shared by languages A and B, occurred earlier in time than change 2, shared by B and C, while in the righthand diagram the order of the changes is reversed. The branching diagrams reflect this difference in the histories of the three languages. However, since the outcome at $time_3$ is identical in both cases, it is not possible on the basis of this final stage to determine which of the two changes took place first, or whether both changes spread slowly, overlapping in time. If there are criteria for determining the sequence in which changes occurred, then it is possible to trace the history of closer relationships among the different groups within the proto-speech community. See Chapter 5 for an example of subgrouping involving changes of the type illustrated schematically here.

1.22C. Non-genetic contact (diffusion, borrowing)

As noted earlier in the case of Brahui, inferences about genetic relationships depend on the assumption that it is possible to differentiate between lexical resemblances resulting from linguistic borrowing (diffusion) and those which reflect the original identity of two languages. For Brahui (see **1.21A**), the proof of relationship to Dravidian rests on the assumption that the resemblances between Brahui and other Dravidian languages represent the oldest segment of the language's vocabulary, while the resemblances to Indo-Aryan and Iranian languages represent subsequent borrowings from contiguous languages. In other cases, the borrowing of words may be the only evidence for prehistoric contact between groups, as in the case of OIA loanwords in ancient Hurrian texts of the second millennium BCE (see **1.32D**). See also the discussion of loanwords in the Rigveda in **3.2**.

Diffusion can take place purely in spoken forms (e.g. from Sindhi to Brahui, or from ancient Dravidian to Old Indo-Aryan), or may be mediated by written language, as is the case with many Latin and French words in English, and many words derived from Sanskrit in modern South Asian languages. Where diffusion is purely through the spoken language, we infer that there was social contact between two groups of a sufficient extent to produce individuals who were (to some extent) bilingual, who thus served as the agents of diffusion.[25] In other words, linguistic diffusion provides evidence of *acculturation* between different sociolinguistic groups.

The process of diffusion requires the intermediation of bilingual speakers, though the bilingualism may be only partial. Uriel Weinreich, in his classic study of bilingualism (1974), introduced the term (*linguistic*) *interference* to denote the process of deviation from the structural norms of a language manifested in the speech of bilingual persons as a result of their familiarity with another language (1974: 1). Weinreich distinguished three types of interference:

(1) *Phonic interference*: For example, failure of native Spanish or Italian speakers to make the English **ɪ–i** distinction (as in **bit-beet**, **sit-seat**), failure of

European language speakers to make the Hindi–Urdu **p-ph-b** distinction (as in **pal** 'moment', **phal** 'fruit', **bal** 'force').

(2) *Grammatical interference*: For example, German–English **He comes tomorrow home** (word order), Hindi–English **Since how long you are working here?** (= 'How long have you been working…?'); Tamil English **Triplicane going, that bus take** 'Take the bus that goes to Triplicane'.

(3) *Lexical interference*: For example, English **thug**, **loot** (from Hindi–Urdu **ṭhag**, **lūṭ)**, American Italian **sanemagogna** 'son of a gun' (Weinreich 1974: 47), German **Wolkenkratzer** (loan-transfer modeled after English **skyscraper**, Weinreich 1974: 51), Pennsylvania German **was-ewer** 'whatever'.[26]

Interference often manifests itself most strikingly in code-switching, a style of speech which alternates between two (or more) different languages: for example,

> **avaṟkku raṇṭu pēṟkkum, they can speak Tamil** 'Both of them, they…' (specimen from an interview with a Tamil–English bilingual), **Mother learned Hindi because bōmbēyil āyirunnu ādyam** '… because she was in Bombay previously' (Malayalam–English bilingual; specimens from Southworth 1974, see **4.31** for further examples).

Apart from the act of code-switching itself, all the types of interference discussed by Weinreich appear with high frequency in this style of speaking, possibly higher than in the ordinary speech of bilinguals.

While it is possible to identify many types of interference, there are cases which are not as clear. For example, Silva-Corvalán, in a study of Spanish–American bilingualism (1983), concludes that changes in Spanish observable in her data (uses of the verb **estar** in place of **ser** in usual native Spanish) were probably accelerated by the speakers' knowledge of English – though the gradual replacement of **ser** by **estar** has been going on independently in Spanish for a long period of time. This raises an important methodological question regarding the status of many linguistic changes which have taken place in the past: evidence of interference in the past must be treated with great sensitivity, with careful attention given to other evidence of contact, both linguistic and cultural. (See further discussion in **3.3**.)

1.3. Linguistic palaeontology

This subfield, also sometimes known as comparative philology, uses the history of the forms of language (attested or reconstructed), along with textual and other evidence, to draw inferences about the socio-historical realities of the ancient societies in which the languages were used.

1.31. Assumptions

The central assumption of linguistic palaeontology is that *words and their meanings can be reconstructed* (**1.22D**), with the corollary that from these words direct inferences can be made about the nature of the ancient speech communities which used the words. Reconstructions are based on *shared forms*: thus, Tamil (South Dravidian) **ā** 'cow' and Kudux (North Dravidian) **ōy** 'cow' lead to the reconstruction of a Proto-Dravidian word ***āy** 'cow' and to the inference that speakers of Proto-Dravidian had some familiarity with cows. The general reliability of the method can be tested where the ancient parent of a group of related languages is known: thus, for example, a group of words in Romance languages including Sardinian **iscriere**, Rumanian **scriere**, Italian **(i)scrivere**, French **écrire**, and Spanish **escribir** 'write' leads to the reconstruction of a Proto-Romance ***skrībere**, which matches Latin **scrībere**.

This basic assumption, and the reliability of the methods and conclusions of linguistic palaeontology, have been challenged by linguists and others. The following quote from a linguist, Ernst Pulgram, is illustrative of these criticisms:

> If we reconstructed Latin on the evidence of the Romanic languages alone, ignoring and neglecting the existence of Greek, Keltic, Germanic and the other ancient Indo-European dialects, and if thereupon we derived from the state of the common Romanic vocabulary conclusions on the culture of the speakers of Latin . . . we might well arrive at the following results: Proto-Romanic **regem** and **imperatorem** show us that the Latins lived in a monarchy under kings or emperors (but what shall we make of **rem publicam** which could presuppose a Latin republic?); since all Romanic languages contain words cognate with French **prêtre** and **évêque**, 'priest' and 'bishop', the Latins were Christians; also words cognate with French **bière**, **tabac**, and **café** are common Romanic, evoking a picture of Caesar's soldiers guzzling beer and smoking cigars in sidewalk cafés; and since all Romanic languages name a certain animal **cheval**, **caballo**, **cal**, etc., and have words for 'war' like **guerre**, **guerra**, the Latins called the horse **caballum** and the war **guerram** and were no doubt warlike people with a strong cavalry.
>
> (Pulgram 1958: 146–7, cited in Renfrew 1987: 85; reprinted by permission of the publishers from *The Tongues of Italy: Prehistory and History* by Ernst Pulgram, Cambridge, MA: Harvard University Press, Copyright © 1958 by the President and Fellows of Harvard College, Copyright © renewed 1986 by Ernst Pulgram)

This passage has been presented by the archaeologist Colin Renfrew with the comment "this marvelous piece of nonsense reconstruction brilliantly exemplifies the dangers of linguistic palaeontology" (1987: 85–6).

Pulgram's point is valid, though it does not justify the conclusion that the whole enterprise is without merit. Practitioners of linguistic palaeontology do not create reconstructions in a vacuum, ignoring the possible influence of surrounding languages. And even when little or nothing is known of the languages contemporary with the period of reconstruction, there are often phonological or semantic clues to indicate that a word may be from another source. For example, the early Indo-European words for 'hemp' (Gk. **kánnabis**, etc.) are suspected, for phonological and other reasons, to be very early borrowings, possibly in Proto-Indo-European itself. See **8.23**, however, for a case in which internal borrowing may create a similar effect.

The uncertainty in reconstructions arises from the state of affairs discussed in **1.22B**: since languages do not always neatly split off from each other, innovations can diffuse from one area to another, even between languages which are already substantially different from each other. This is the explanation, for example, of the presence of words for 'tobacco' in the Romance languages. And though the painstaking linguistic palaeontologist may often find clues to the history of such words, there is no guarantee that *all* such later diffusions can be identified. Therefore a certain degree of indeterminacy must be accepted regarding the *contemporaneity* of the different reconstructions which make up a proto-language. Though we treat proto-languages as if they were real languages, some vocabulary items may belong to the latest period of linguistic unity among the daughter languages, while others represent innovations which diffused among the separate languages at a later time. To this extent, then, a proto-language is likely to be a composite of features (phonological, lexical, and grammatical) which never existed all at once in the same time and place: we cannot guarantee its chronological homogeneity.[27] Such reconstructed languages can have their uses, especially if we are aware of this indeterminacy. And by careful examination of archaeological contexts in relation to our reconstructions, we can perhaps quantify this indeterminacy to some extent (see **8.4**).

It is also true that some reconstructions are more reliable than others. Generally speaking, comparative reconstruction works best when we can count on at least two independent witnesses, that is, languages or language groups which show little possibility of recent diffusion, either because of long geographical separation or for other reasons (such as extreme sociolinguistic conservatism, as found for example in Lithuanian among Indo-European languages, and in Tamil among Dravidian languages). So, for example, the most reliable Proto-Dravidian words are those which are found in North and South Dravidian, two branches that have not had direct contact for a millennium or more.[28] (See further discussion of the Dravidian case in **8.2**.) The following two sections discuss additional ground rules regarding reconstructions.

1.31A. Meanings of proto-forms

The meaning assigned to a reconstructed form can be no more specific than that meaning which is *shared by all the cognate forms*. For example, OIA **yava** means

'barley', but since the other Indo-European cognates refer either to 'grain' in general or to other (specified or unspecified) grains, the only meaning which can be assigned to the Proto-Indo-European form is 'grain' or 'a grain' (see M53, M86 s.v.). A South Asian example involves a set of words which mostly refer to rice in modern Dravidian languages, but can only be reconstructed in the meaning 'seed' in Proto-Dravidian (Chapter 8, Appendix A, D1).

1.31B. Interpreting the absence of proto-forms

The absence of a reconstructible term for a certain artifact, cultigen, natural feature, or the like in no way proves the absence of the object in question in the proto-culture. It simply means that no word can be reconstructed for that item. The item might well have existed, and have been designated by a single word in the proto-speech community, but if that word was replaced by another (or if the object or concept referred to by the word went out of use as a result of culture change) in one or more branches, it cannot be reconstructed. For example, contrary to the statement by Pulgram quoted in **1.31**, no word can be reconstructed in Proto-Romance for 'war', since the Latin **bellum** 'war' was replaced in the western Romance languages by a Germanic word **werra**, which Hall reconstructs as Proto-Italo-Western Romance ***gwerra** (Hall 1976: 203): cf. French **guerre**, Spanish **guerra**, Italian **guerra**, etc. Without the testimony of Latin, one might be tempted to wonder whether the speakers of Proto-Romance knew warfare. For a South Asian example: the OIA words **nagara** (see **3.22A(3)**) and **pattana/paṭṭana** (**3.22A(4)**) both mean 'town', and both occur in works belonging to the late Vedic period (*c*.1250–750 BCE). It is generally agreed that these words probably came from Dravidian (see **8.31** for details). However, these words can only be reconstructed to a relatively late level in Dravidian, because they have been replaced by borrowings from Indo-Aryan in all the Central and North Dravidian languages, which are surrounded nowadays by Indo-Aryan languages. Thus it is probable that these words existed in early Dravidian, from which they were borrowed into OIA, and were subsequently replaced in the northern and central branches of Dravidian.

1.32. Inferences in linguistic palaeontology

Among the areas of prehistory commonly dealt with in linguistic palaeontology are material culture, socio-political structure, religion, habitat, migration, and interethnic contacts. Inferences on many aspects of culture can be drawn from reconstructed vocabulary, following the principles noted earlier.

1.32A. Inferences about material culture

Chapter 8 provides an example of the use of reconstructed vocabulary to draw inferences about material culture, in this case that of Proto-Dravidian. For earlier

work on Dravidian, see Sjoberg 1971a,b (with references). See Rau 1971, 1972, 1974, 1983 for work on OIA; Thapar 1984 for Indo-Aryan socio-political structure. See Zide and Zide (1973), Zide and Zide (1976) for Munda agriculture.

1.32B. Inferences about prehistoric habitat and migrations

The problem of the Indo-European "homeland" has been approached by a number of scholars using the evidence of reconstructed vocabulary: see for example Friedrich (1970, with references). In Chapter 7, evidence of plant names is used as an indication of the likelihood of the earlier presence of Dravidian speakers in the northern Deccan. The study of place names can often provide more explicit evidence of the earlier presence of a particular language in an area. See Chapter 9 for a discussion of Dravidian elements in the place names of Maharashtra and other parts of South Asia.

1.32C. Inferences about the nature of prehistoric ethnic contacts

Evidence of prehistoric borrowing of words from one language to another leads not only to an inference of prehistoric contact and acculturation between groups (**1.22C**), but often provides explicit clues about the cultural areas of contact, and in some cases the chronology of contact. (See Chapter 3 and **10.21A**.)

1.32D. Chronology

Sequences of interrelated linguistic changes can provide information on the ***relative chronology*** of linguistic stages. For example, Krishnamurti (1969), quoting Burrow, notes that the change Proto-Dravidian ***c** → South Dravidian **∅** (as in Tamil **akkuḷ** 'armpit': Telugu **cakkili** ← PD ***cakk…**DEDR2274, Tamil **āṟu** 'six': Naiki **sādi** ← ***cāṟ…**DEDR2485) must have been completed before the change ***k** → SD **c/#__i** (see **1.21B**), since there are no cases of **∅** ← **c** ← ***k**.

Such *relative* chronologies can sometimes be converted into *absolute* chronologies on the basis of dateable texts. In this case, it is possible to date these changes to a period intermediate between the third century BCE and the time of the earliest Tamil classics, on the basis of words found in texts (Krishnamurti 1969: 317). Another example is provided by the OIA words occurring in Hurrian texts of the fourteenth to fifteenth centuries BCE, which show that the separation of Old Indo-Aryan from Old Iranian must be prior to that period (Mayrhofer 1974). See **1.5** for further discussion of methods of measuring chronology.

1.4. Sociolinguistics

Making inferences from linguistic form to social context, or the reverse, requires some general notions of the possible relationships between language and social situations. Recent work in sociolinguistics and in the ethnography of communication

has contributed many insights into these relationships as far as the contemporary scene is concerned (see Hymes 1964; Gumperz and Hymes 1972; Labov 1972, 1980, 2001; Bauman and Sherzer 1974; Gumperz 1982a,b among others). While much more work is needed on the sociolinguistics of pre-industrial societies, sociolinguistic research has nonetheless provided some indispensable tools for the general understanding of the social context of linguistic change.

1.41. Assumptions

The assumptions of sociolinguistics which are most relevant for historical linguistics concern the following aspects of relationships between language and society: (A) the presence of variation in all spoken languages; (B) the social functions of linguistic variation; (C) the role of linguistic variation in sociolinguistic change; (D) relationships between acculturation and linguistic diffusion (borrowing).

1.41A. Variation

Briefly, sociolinguistic studies have shown that there is no spoken language without variation. While there is as yet no way to predict what features will vary, it is clear that all subsystems of a language (phonology, syntax, and semantics) are potentially subject to variation. Following are the most common types of variation:

(1) phonological variation, for example presence or absence of **r** in words like **guard**, **far** in New York City English;
(2) morphological variation, for example presence or absence of a past tense marker in words like **missed**, **asked** in various American dialects;
(3) syntactic variation, for example **She drinks coffee in the morning** vs **In the morning she drinks coffee** (the latter being more frequent among lower-class American speakers according to Callary 1971);
(4) discourse variation, for example the use of generalizing phrases such as **and things like that**, found more frequently among working-class than middle-class speakers in Dines's (1980) study of Australian usage.

In terms of sociolinguistic distribution, there are three basic types of variation:

(1) regional variation, for example the difference between the eastern part of the United States which distinguishes **don** from **dawn** and **cot** from **caught**, and most of the central part of the country which does not;
(2) social or stratificational variation, for example middle class **didn't see any** vs working class **didn't see none** in many parts of the United States;[29]
(3) contextual or stylistic variation (when the same individuals use different variants depending on the social situation), for example **What are you doing?** vs informal **Whatcha doin'?** in American English.

William Labov and others have pointed to the existence of *stable sociolinguistic variables*, which exist in all modern societies that have been studied by sociolinguists. A well-known example in English (both in the US and Britain) is the (**ing**) variable, that is, the variation between **going**, **singing**, **fishing**, etc. and **goin'**, **singin', fishin'**. Others are the variation between stops (**t d**) and fricatives (**th dh**) for the initial sounds in **thin** and **then**, and the presence or absence of consonantal **/r/** in words like **guard** and **car**, in New York City. Figure 1.3 is a schematic depiction of this type of variable, with the different lines in the diagram depicting different socioeconomic classes, however defined (starting with the lowest class at the top of the figure) and the vertical scale representing different styles of speech, from the most casual or spontaneous on the left to the most careful or formal on the right. The vertical scale shows the proportions of usage of the variant forms, with higher numbers indicating greater frequency of the non-standard or less formal variant. The slope of these lines is typical: all classes are ranked in the same order for each style of speech, and speakers of each class differ in the same direction in their use of the variants. Thus the variation is simultaneously responsive to the interpersonal context of speech, and to the social status of speakers. Such stable variation may continue for generations, or even centuries; the (**ing**) variable seems to go back to Old English (Labov 2001: 418–19).[30]

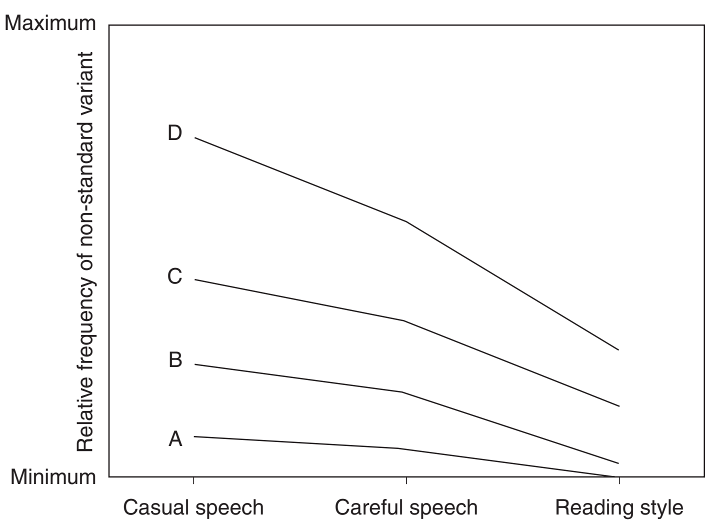


Figure 1.3 Schematic representation of stable sociolinguistic variation (after Labov 2001: 81ff.).

Notes
This can be viewed as a generic diagram, showing the typical relationship between classes and speech styles in stable sociolinguistic variation. The lines A, B, C, D depict different socioeconomic classes, however defined (A = highest). The horizontal scale represents different styles of speech, from the most casual or spontaneous on the left to the most careful or formal on the right. The vertical scale shows the proportions of usage of the variant forms, with higher numbers indicating greater frequency of the non-standard or less formal variant.

1.41B. Social functions of variation

The importance of linguistic variation in society is related to the various *roles* of language in people's lives. Language transmits not only *content* (e.g. "I'll pick you up at eleven-thirty at the corner of Forty-eighth and Lombard", or "Gimme a large pizza with extra cheese and three medium Cokes"), but also the *social identity* of the speaker/sender (indicated for example by regional forms like **any road** vs **anyhow**, or socially significant variants like **He doesn't have any** vs **He don't have none**), as well as the *role relationships* the speaker is manifesting (indicated by the choice of formal or informal variants, as well as by specific words or phrases such as **gimme**, **please**, **I'd like to…**, **Would you mind…**, etc.). The stable sociolinguistic variables like (**ing**) mentioned before form part of the code by which people communicate their places in the social structure and the roles they are performing. Some might say that a language without this kind of variation would be dysfunctional, since its speakers would need to find other means to express these pervasive distinctions.

1.41C. Variation and linguistic change

However, while variation is ubiquitous in spoken language, not all differences "make a difference" in terms of linguistic change. Though variations may occur as a characteristic of individuals or small groups without any social significance, *change* begins to take place primarily when these variations are perceived (consciously or unconsciously) as being characteristic of a particular context (social, regional, or stylistic), and are accordingly imitated or avoided – again, this action is often unconscious – by some individuals or groups (see Labov 2001: 503–6). For example, the loss or attenuation of English **r** was initially a matter of regional variation without any particular social significance, which only later came to be an important status indicator (see McDavid 1948; Labov 1972: ch. 6).

The existence of stable sociolinguistic variation (**1.41A**) demonstrates the relevance of social structure and social context to language variation and change. However, it is important to recognize that there is a crucial difference between social dialects and regional dialects. While it is possible to associate regional linguistic variants with the populations of specific areas, the same is not true of social variation. To begin with, a social class is an abstraction which, strictly speaking, has no members, since it is based on a constellation of features which can be statistically manipulated as needed; furthermore, most of the variables used to define social class are continuous in nature, without discrete boundaries. Though we may be required to divide our population into three or four (or more) classes in order to display certain sociolinguistic relationships clearly (as in Figure 1.3), such figures depict *patterns* or *relationships* rather than discrete varieties of speech. Such diagrams generally show a polarity between two variables, one usually regarded as correct (formal, polite, educated, proper, or simply *good*) and the other as incorrect (informal, impolite, uneducated, improper, or just plain *bad*).

This polarity, which is often lacking in studies of regional dialects, is *relative to the context of usage*; as indicated by the schematic Figure 1.3, most people's usage is somewhere between the extremes, and varies from context to context: the speech of a particular individual on a particular occasion can be ranked relative to the speech of other individuals, or relative to the same individual's speech on another occasion, with respect to particular variables, but not classified as belonging to a particular dialect. And since relative frequencies are involved, it is never possible to predict which variant an individual speaker will use in a specific utterance.[31]

Sociolinguistic studies of linguistic changes in progress have shown that changes usually originate in socioeconomic groups that are non-peripheral (i.e. neither the highest nor lowest groups), from which they spread to the limits of the local community, and sometimes beyond (Labov 2001: 500ff.). In a number of studies carried out in cities in the United States, those social segments commonly referred to as the "upper working class" and "lower middle class" (which of course are not distinct classes, but part of a multi-dimensional continuum) have been found to be most active in initiating and propagating changes. Though the changed forms are often stigmatized by authorities in education and the media, in many cases they enter the usage of all members of the local community, including the upper class (variably, that is more in some spoken styles than in others), thus supporting the assumption made above that the linguistic usage of members of any group tends toward convergence.

As the result of a search for a more specific identification of the agents of linguistic change, Labov's long-term study of Philadelphia speech came to the following conclusions:

(1) ongoing linguistic changes reflect nonconformity to established social norms, and arise "in the social milieu that most commonly defies those norms";
(2) linguistic changes "are generalized to the wider community by those who display the symbols of nonconformity in a larger pattern of upward social mobility" (2001: 516).

In other words, linguistic change begins as a form of social protest, directed against the upper levels of the existing social hierarchy, as a way of expressing rejection of mainstream values and asserting the validity of one's own; it is perpetuated by those who, while adopting some of the major values of the establishment, continue to assert their separateness from it. The finding that "leaders of linguistic change" can be identified in terms of their social characteristics is important – and perhaps surprising – information, strongly supported by a wealth of empirical data collected over a long period of time. Whether these conclusions will be found to be valid for other societies is an open question; see further comment below.

Several other important social factors emerge from studies of this kind. One is the role of women: Labov found that in Philadelphia, the large majority of the

ongoing changes were both initiated and advanced by female speakers (2001: chs 11–12), and the same was true in a study carried out in Cairo, Egypt by Haeri (Haeri 1996; Labov 2001: 408). While Egypt is mainly a Muslim country, its administration has been determinedly secular, and thus there are fewer restrictions on women's participation in public life than in some other Muslim-dominated polities. Looking at the question of gender in terms of linguistic change in earlier periods, there are several factors that would possibly differ considerably from the situation in present-day Philadelphia or Cairo.[32] In many parts of South Asia, the public behavior of females is strictly controlled from puberty onward.

The implications of this for language can be seen in Gumperz's study of phonological variation in a North Indian village (Gumperz 1958), which concluded that the variation could best be accounted for in terms of male friendship groups which served as the foci of most social interaction at the village level. In this context women would not be expected to play much of a role in language change, since many village brides were brought in from distant villages, often located in quite different dialect areas, and were generally confined to their husbands' houses under the thumbs of their mothers-in-law. Even older women who had greater freedom of movement seldom left the village, which was mainly a male prerogative. In my own study of lexical variation in Tamil (Southworth 1979b), it was clear that the language of rural women tended to be restricted to the usage of their own locality and caste, whereas men generally had wider exposure to the usage of other regions and other social groups.[33]

Labov also notes that African Americans in Philadelphia do not participate in the ongoing changes which were found in the speech of all white Philadelphians, and in fact this seems to be generally true of all those classed as "non-white" (Black, Hispanic, Native American) in many US cities (Labov 2001: 506–8). Labov attributes this to the "large and increasing residential segregation of African Americans in the Northern Cities," and notes that the same barriers are not found in European countries. This situation can perhaps be likened to the position of castes in traditional South Asia, especially to the boundary between the lowest castes (formerly untouchables) and the rest. Gumperz's study of intravillage variation (cited earlier) noted that the speech of members of the lowest castes differed most from those of other villagers, while the speech of most upper-caste villagers differed in only minor ways. In fact, the three lowest castes in the village each lived in separate quarters and differed in their speech from each other as well.

In the study of lexical variation in Tamil, many speakers were unaware of differences in the usage of members of other castes. And while residential segregation no doubt accounts for part of these differences, several commentators on the South Asian scene (e.g. Pandit 1979) have noted that *type* of communication may be as important as *frequency*: an untouchable worker might spend most of the daylight hours working in the fields with members of other castes without this having much effect on his or her linguistic usage, because of the importance of one's home group (caste and extended kin) in determining speech habits,[34] and

the associated taboo on the imitation of high-caste behavior by members of the lower castes.

It is not yet clear how the findings of studies of change in the twentieth century can help us to understand linguistic change in the distant past, inasmuch as these findings refer to constructs such as socioeconomic class, upward mobility, and social norms which are transmitted by (among other factors) formal education. Labov acknowledges this problem when he says, with regard to his five-step scheme of transmission of change:

> One might call these Principles of *Urban* Transmission because the social patterns described are typical of the social stratification of large cities and the operation of the socioeconomic hierarchy . . . reservations must be made in regard to more remote societies and more remote periods of history. The mechanism . . . is based on a high degree of social mobility. In those societies where class stratification takes different forms . . . we must be ready to modify the uniformitarian principle in favor of a more historically specific account of the mechanism of change.
>
> (Labov 2001: 437, reprinted with permission from William Labov, *Principles of Linguistic Change, Vol. 2: Social Factors*, Malden, MA and Oxford: Blackwell Publishers Ltd, Copyright © William Labov 2001)

Social hierarchy has been part of human history for a long time: it existed in the early cities of the Indus Valley in the third millennium BCE, and there is evidence for it in the earliest reconstructible form of Dravidian, probably dateable to the same period (see **8.38**). And yet, even in contemporary South Asia there are Munda-speaking groups like the Savara which still retain an egalitarian way of life with very little formal authority structure (David Stampe p.c. 2003). Since the Munda languages have undergone linguistic changes which do not seem to be different in kind from those occurring in other language families, it will be necessary to look for other kinds of social divisions which might have served the same function as socioeconomic class – and/or other kinds of social dynamic which would have the same effects as those described for modern American cities. Of course, the population sizes of these early groups were presumably much smaller than those of the twentieth century. The few studies of ongoing linguistic change which have been carried out on small pre-industrial societies (e.g. Moylan 1981, 1982) suggests that here, smaller-scale processes like marriage patterns and residence rules may be primary factors in language change, rather than the large-scale class and ethnic factors operating in large industrial societies.

On the other hand, Labov's statement that "social structure affects linguistic output through changes in the frequency of interaction" (2001: 506) is probably applicable to many cases of linguistic change, regardless of social type or historical period. In looking at the problem of language change in a long-term

perspective, it will be extremely useful to identify the kinds of social changes that might lead to such changes in patterns of social interaction. There are some hints to be found in modern sociolinguistic studies. Labov notes, for example, that in Philadelphia "the oldest working class neighborhoods are the geographical and cultural matrix in which the local dialect was formed and continues to advance" (2001: 502), pointing to the potential importance of settlement patterns as a factor in linguistic change. Elsewhere Labov speaks of immigration and its effects on change (2001: 503–4), and at several points suggests that the social disruptions caused by First and Second World Wars (e.g. 2001: 227–8, 315) may have led to alterations in interaction patterns which stimulated linguistic changes. Presumably, any large-scale changes in the composition of the speech community (or in small societies, even small-scale changes) have the potential of triggering language change. The task of correlating linguistic and social changes in a cross-cultural and historical framework is vital for linguistic archaeology, and work on this task has barely started.

1.42. Inferences from sociolinguistics

The relevance of sociolinguistics to linguistic prehistory is manifest with regard to two areas: (A) the social background of linguistic change, and (B) the social contexts of linguistic convergence.

1.42A. The social background of linguistic change

Labov's linguistic study of Martha's Vineyard (Labov 1963) was the first clear documentation of the relationship between pronunciation and social identity – hence the appropriateness of its title: "The social motivation of a sound change." Assuming that similar processes have operated in the past, I suggest in chapter **6.14B** that the change **r** → **l** in MIA may have correlated with a certain anti-establishment and anti-Brahmanical ethos which crystallized around the early adherents of Buddhism (see **6.16**).

1.42B. The social contexts of linguistic convergence

It has long been obvious that when new words enter one language from another, we can infer that some kind of social contact has occurred between speakers of the two languages. However, few attempts have been made to discover which specific *kinds* of social contact can be inferred from particular configurations of borrowed elements. Chapter 4 explores the possibilities of establishing correlations between types of borrowed elements (e.g. different parts of speech, grammatical morphemes, affixes, grammatical rules) and different kinds of contact – in terms of the *intensity* of contact (i.e. the closeness of social integration of the groups) and the *range* of contact (the extent to which different segments of the society are affected by the contact).

1.5. Glottochronology

The technique of glottochronology (lexicostatistics) has been thought by some to provide reliable, though very approximate, estimates of the chronology of prehistoric separation of related languages (see Swadesh 1951, 1952; Gudschinsky 1956). Apart from the widespread skepticism which the theory has encountered, the results have not been adequately tested in cases where the chronology is known from other sources. Furthermore, there have been numerous careless applications of the method (see Krishnamurti 1980). Nevertheless, the method may be of some help when no other information is available, if only to show degrees of relatedness within a group of languages. See **5.24** for an example.

1.6. Philology

The field of philology developed from the study of classical texts, and became more or less synonymous with (historical) linguistics in the nineteenth and early twentieth centuries. As opposed to the broader field of linguistic palaeontology (**1.4**), philological study *per se* is concerned with the interpretation of ancient texts, and also with investigating their authenticity and establishing their original form to the extent possible. While philology is now considered to be distinct from linguistics, it is of course an important part of historical linguistics, since linguists – if they themselves are not also philologists – must depend on philologists to provide authenticated texts as a basis for their analyses. Because of the existence of many ancient texts in Old and Middle Indo-Aryan, Old Iranian, Old Tamil, and other Dravidian languages, philology is clearly crucial for the historical linguist working in South Asia. Philological studies of the Rigveda and the other Vedas, for example, have not only helped to clarify the meanings of words and passages in the texts, but have also been able to locate the earliest forms of these texts in time and place, on the basis of both internal and external evidence. For examples of philological contributions to the study of OIA see works by Michael Witzel and Wilhelm Rau cited in the bibliography.

1.7. Correlating linguistic and archaeological evidence

The field of linguistic archaeology, as defined in **1.1**, is an approach to prehistory which draws on a number of subfields of linguistics, and makes no pretensions to being a kind of archaeology. We can hope, however, that by correlating the findings of linguistic archaeology with those of (real) archaeologists, a fuller picture of prehistoric societies can be drawn. Before beginning this discussion, it should be pointed out that in attempting to correlate work in two different fields, it is important to ensure that the conclusions of each field are arrived at independently of the other. Colin Renfrew has pointed out that in some cases, archaeologists have made use of the conclusions of historical linguists which were based on (unstated) assumptions borrowed from earlier archaeology – assumptions which

would have been instantly rejected by modern archaeologists if they were made explicit (Renfrew 1987: 18). In order to avoid such a situation, it is important to ensure that each body of evidence stands on its own. Therefore the prehistoric inferences made in the following chapters will be based on linguistic evidence, and connections with archaeology will be explored only after these linguistic inferences have been made explicit. (See especially Chapters 6–9 and the concluding discussion in **10.3**.)

The bulk of this chapter has focused on the kinds of evidence used by linguistic archaeologists, and the kinds of prehistoric inferences which this evidence has led them to. In most cases, however, these linguistic inferences have remained nothing more than inferences, unsupported by any kind of data from the real world. For example, attempts to identify particular archaeological complexes with particular languages have not met with wide success – in part because there have often been competing, mutually exclusive hypotheses. The following quote from an archaeologist, Lamberg-Karlovsky, illustrates the problem:

> Russian scholars working in the Eurasiatic steppes are nearly unanimous in their belief that the Andronovo culture and its variant expressions are Indo-Iranian. Similarly, Russian and Central Asian scholars working on the Bactrian Margiana complex share the conviction that it is Indo-Iranian. The two cultures are contemporary but very different. Passages from the Avesta and Rigveda are quoted by various researchers to support the Indo-Iranian identity of both, but these passages are sufficiently general as to permit the Plains Indians an Indo-Iranian identity. Ethnicity is permeable and multi-dimensional, and the "ethnic indicators" employed by Kuzmina can be used to identify the Arab, the Turk, and the Iranian, three distinctive ethnic and linguistic groups. Ethnicity and language are not so easily linked with an archaeological signature.
>
> (Lamberg-Karlovsky 2003: 74; reprinted with permission from C. Lamberg-Karlovsky, 'Archaeology and language: the Indo-Iranians', *Current Anthropology* **43**: 63–88, Copyright © 2002 by the Wenner-Gren Foundation for Anthropological Research)

It also seems to be true that archaeological work can proceed quite satisfactorily without knowledge of the ethnicity and/or language of the people who occupied the sites under study. The archaeological study of Mohenjodaro and Harappa has not been hampered by the lack of progress in the decipherment of the Indus script – though of course when such information is available, as in the case of many sites in Mesopotamia, Egypt, Central America, and elsewhere, it is welcome. Thus it may be difficult for the linguist to persuade his archaeological colleagues that historical linguistics is even relevant to archaeology. Presumably knowledge of prehistoric languages would be relevant only to the extent that such knowledge helps to answer archaeological questions. And the knowledge (or inferences)

offered by linguistics cannot substitute, of course, for contemporary texts or inscriptions which can reveal what a particular artifact or building was used for, or which gods were worshiped at a particular temple, etc. And yet language is an important part of social behavior: it is the glue that holds a society together and allows the perpetuation of culture, even in cases where some members of a society speak more than one language.

Historical linguists and other prehistorians can be said to study different aspects of the same reality, and yet our conclusions are not often of much relevance to each other. I would argue that the question of relevance can be resolved by recognizing that each discipline has a grasp on only part of the ancient reality. However, it was an archaeologist, Colin Renfrew, who first put this issue in its proper perspective. His words on the subject are worth quoting at length:

> I have tried to show . . . that there has hitherto been no valid methodology for matching the evidence . . . from historical linguistics with the material evidence from archaeology . . . the primary difficulty in the entire enterprise is that the archaeological evidence from an early (and non-literate) period can tell us nothing directly about the languages which were spoken . . . At least the archaeological remains, the material culture, can be set firmly within a chronological framework . . . The discipline of historical linguistics suffers from the disadvantage that, even when it is possible within reasonable plausibility to construct an early language from more recent evidence, there is no way of setting that firmly within a chronological framework. . . .
>
> But ‘when?’ and ‘where?’ are precisely the questions which archaeologists in their prosaic way like to ask, and are equipped to answer. The real problem is to bring these two worlds of argument . . . into some sort of constructive relationship . . . The task is to be undertaken by understanding better the relationships between the *processes of change*: on the one hand the linguistic process, on the other hand those changes which leave material traces in the archaeological record. The mediating phenomena appear to be largely social and demographic. Linguistic change does not take place in a vacuum, irrespective of other features of society: that is one of the emerging lessons of sociolinguistics. And those factors in society which correlate with or promote linguistic change are at the same time influential in the material sphere and find traces in the archaeological record. This is the nexus which requires further investigation, and where I feel optimistic that progress can be made.
>
> (Renfrew 1987: 285–6, extract from *Archaeology and Language: The Puzzle of Indo-European Origins* by Colin Renfrew published by Jonathan Cape; used by permission of the Random House Group Limited).

Even if archaeologists and linguists can agree with Renfrew's approach – and there is no evidence that his words have had any substantial impact on communication between linguists and archaeologists in the fifteen years since he wrote them – there remain many specific issues to be worked out, involving each field's assumptions and rules of evidence, along with understanding the meanings and implications of specific terms like "archaeological complex" and "archaeological culture," and linguistic terms like "reconstruction," "speech community," "dialect continuum," and even "language" and "dialect." Though there have been any number of joint conferences, it is my contention that, at least in the field of South Asia, on the whole linguists and archaeologists have not really learned to listen to each other. There are a number of reasons for this: even though we work with different aspects of the same reality, we differ in our approaches, our assumptions, our terminology, our methods, and our expected outcomes.[35]

The following formulation of Renfrew's suggested approach may be a useful guide at this point in the discussion:

> In place of the old framework of linkages – specific language ↔ people/ethnos ↔ specific archaeological "culture" – it may be possible to develop…a rather different framework of inference: language change ↔ economic/social/demographic change ↔ change in the archaeological record."
>
> (1987: 141, extract from *Archaeology and Language: The Puzzle of Indo-European Origins* by Colin Renfrew published by Jonathan Cape; used by permission of the Random House Group Limited)

The changes which the historical linguist reconstructs often seem to imply large-scale social changes. As Renfrew says, "The mediating phenomena appear to be largely social and demographic" (1987: 286; see the longer quotation previously).

At this point it is necessary to acknowledge the unfortunate absence of any general theory of the relations between linguistic and social change, or even any comprehensive treatment of the subject. This is a task which properly belongs in the realm of sociolinguistics or the sociology of language (see **1.41C**, end). In the absence of such a body of work, the following remarks will have to suffice for the present purpose. (Section **10.3** takes up this question again, using the findings of Chapters 3–9 as material for discussion.)

The data which provide input for the inferences of historical linguistics consist of

(1) geographical distribution of languages and language features;
(2) changes in linguistic structure;
(3) vocabulary and changes in vocabulary (including loanwords);
(4) evidence of language contact, including language shift and language loss.

Perhaps the first category is the one most often invoked in discussions of language and archaeology. Two groups of people, A and B, living in widely separated areas, are shown to speak related languages. Clearly there is an implication of movement, leading to obvious questions like: 'Which is the original homeland?', 'What prompted some people to move?', and 'How and when did the movement take place?' Migrations can have a number of different causes – such as environmental changes, floods, epidemics, natural disasters – as well as a number of different consequences, such as contact with other groups (including subjugation, warfare, etc.) Many of these processes leave traces in the archaeological record.

Though it has been shown many times that changes in linguistic structure can be tied to social factors such as age and social class, we cannot yet say to what extent it may be possible to correlate these social factors with material evidence. Reconstructed vocabulary, on the other hand, can serve to identify many aspects of a group's material existence, social and political structure, beliefs, relation with the environment, etc., and how these have changed over time. See Chapter 8 for an example of this type of inference, and a suggested correlation with the findings of archaeology (**8.4**).

Evidence of language contact, including loanwords and borrowed structures, plays a large role in linguistic archaeology. Language contact situations also have many different outcomes, from bilingualism or partial bilingualism (including pidgin languages) in zones of contact to much more intimate integration of groups, leading in some cases to large-scale bilingualism, language shift, and language loss. Any of these linguistic events imply large-scale social changes, though they may happen gradually. Often the only evidence of the existence of a prehistoric speech community is the traces that it has left in another community's language (see Chapter 3). Chapter 4 looks at different types of language contact situations, and explores possible correlations between the amount and type of borrowed items, and the kind of contact which took place.

Thus the kinds of social changes which can be inferred from our linguistic evidence include migrations, interethnic contacts – including trade, warfare, colonization, technological changes (in subsistence, metallurgy, irrigation, shipbuilding, etc.), socioeconomic changes, including the change from egalitarianism to hierarchy, changes in relations between people and the environment, etc. As Renfrew proposes, it is these changes themselves, as reflected on the one hand in the archaeological record and on the other in the vocabulary and structure of human language, that offer the possibility of fruitful collaboration between linguistics and archaeology. See **10.3** for a more specific discussion of this issue in the South Asian context, based on the findings of Chapters 3–9.

1.8. Conclusion

While historical linguistics and linguistic palaeontology are well-established subfields of linguistics, there is much room for development in the use of insights

derived from contemporary sociolinguistic research. Our understanding of the relationship between language and society, based on studies of contemporary linguistic communities, has hardly begun to be applied to prehistory. The present work is only able to illustrate a small fraction of the possible applications.

Renfrew's proposal that social changes, as reflected in the archaeological record and in human language, can provide a meeting point for archaeologists and historical linguists, seems not to have received much notice even among archaeologists. The present work is, among other things, an attempt to provide more evidence of the relevance of such an approach, both to linguistics and to archaeology. In spite of some recent work by archaeologists on linguistic questions, as well as many prehistoric hypotheses put forth by historical linguists, the needed dialogue between practitioners of the two disciplines has yet to make a serious beginning.

1.9. Summary

This chapter, intended primarily for non-linguists, discusses the scope of the field of study I have designated as "linguistic archaeology". This field uses the assumptions and methods of five subfields of linguistics to reconstruct a *socio-linguistic context* for ancient linguistic forms, that is, to make inferences from the language of ancient texts and/or linguistic reconstructions about the groups who used the language in question. The five subfields are:

Historical-comparative linguistics (**1.2**), which makes inferences about historical relationships among languages on the basis of shared forms, and which (under favorable conditions) makes possible the reconstruction of earlier stages of languages.

Linguistic palaeontology (**1.3**), which uses the history of the forms of language (attested or reconstructed), along with textual and other evidence, to draw inferences about the socio-historical realities of the ancient societies in which the languages were used.

Sociolinguistics (**1.4**), which provides information about the linguistic and social roles of linguistic variation, and in the present context provides important background for the study of the social contexts of linguistic change, as well as the possible relationships between culture contact and linguistic diffusion (convergence or borrowing).

Glottochronology or lexicostatistics (**1.5**), a technique which may provide useful (though very approximate) estimates of the chronology of prehistoric separation of related languages.

Philology (**1.6**), not strictly a subfield of modern linguistics, is concerned with the interpretation of ancient texts, as well as their history.

The first four of these subfields are each discussed in terms of their basic assumptions and the types of inferences which they generate, with examples from South Asia and other parts of the world.

Linguistic archaeologists also hope to be able to find ways to correlate the linguistic prehistories inferred by these methods with the findings of archaeologists and other prehistorians. Section **1.7** discusses the need for a new kind of dialogue between historical linguists and archaeologists, following a suggestion by Colin Renfrew.

Notes

1 ‘*Linguistic archaeology*’ is clearly a term which could have been independently innovated, perhaps repeatedly. I used this term, along with ‘linguistic stratigraphy’ in works printed in 1974 (see Southworth 1974; Southworth and Apte 1974; Southworth and Daswani 1974). Others have been using the term ‘linguistic archaeology’ since around that time and possibly before, but I do not know of any formal definition of this term before the one given here. The term has also been used by scholars in other fields to designate what I refer to here as *linguistic palaeontology* (**1.3**).

2 Though philology is not recognized as a subfield of linguistics in the context of modern linguistic analysis, it can be so considered in the present context. See **1.6** for further discussion.

3 The distinction between different languages and different dialects cannot be made on the basis of objective linguistic criteria. In real life political, historical, and other factors enter in. Thus, mutually unintelligible spoken varieties of Chinese are treated as dialects of the same language because they share a single writing system and belong to the same political entity and share a cultural tradition, while Hindi and Urdu, which are mutually intelligible in many colloquial forms, are treated as distinct languages using the same criteria (different scripts, different political entities, partly distinct cultural traditions).

4 In practice, many items are of unknown origin; see **3.23**. And in many other cases the best we can do is suggest the highest probability. See **3.23** for copious examples.

5 Reconstructed forms are conventionally preceded by asterisks to denote non-attestation.

6 But note a similar innovation in Hindi–Urdu **do-n-õ** ‘both’, on the analogy of **tīn-õ** ‘all three’.

7 The same applies to borrowed forms which have become integrated into the language, that is, when they are no longer regarded by speakers as foreign but are accepted as normal words in the language (such as English **loot** from Hindustani, or **moccasin** from Amerind). Initially, borrowings enter the language through the speech of bilinguals or partial bilinguals, and often resist native inflectional and derivational processes until they have passed into the speech of monolinguals, a process which may take a generation or more. (See Chapter **4.3** for further discussion.)

8 A comprehensive discussion of types of language change can be found in Polomé 1990.

9 For example, Marathi **/ċ/** (alveolar) and **/ć/** (palatal) were at one time positional variants of the same phonological element, with the latter occurring only before **/i e y/**, but later came into contrast because of (a) phonological changes, including the loss of the penultimate syllable in words like **tyāćiyā** ‘his’, **bhāvāćiyā** ‘brother’s’ (before feminine plural nouns), resulting in the new forms **tyāćā, bhāvāćā** which contrast with **tyāċā, bhāvāċā** (before masculine singular nouns), and (b) the introduction of words with **/ć/** from other languages such as **ćandra** ‘moon’, **ćatūrthi** ‘4th lunar day’.

10 From the Anglo-Saxon Chronicle for the year 994, quoted in Algeo (1972: 135).

11 An Indian acquaintance of mine learned certain English expressions while living with an American family (such as **bumpety-bump traffic** for ‘bumper-to-bumper traffic’), and only later came to know that these were not generally used in American English.

12 For example, Marathi **bahin** and H-U **bahan** 'sister' appear to be derived from an unattested OIA or MIA ***baghini**, with an irregular metathesis of **h** in comparison with the attested OIA **bhaginī**.

13 The aspirate consonants (**kh, ph, dh**, etc.), though transcribed with two symbols, are considered single phonological units, not consonant clusters.

14 As in many studies of linguistic change in progress, the changes in question turn out to involve *linguistic variables*, which appear to varying extent in the speech of many, if not all, individuals in the community, and show statistical correlations with a number of social variables – in this case caste, social class, and education – as well as (socio)linguistic variables such as the formality or informality of the speech situation, the languages or dialects shared by participants, etc.

15 In this type of notation, the symbols following / indicate the linguistic environment in which the change takes place: in this case # indicates word beginning. In general, the formula A → B/X__Y means 'A changes to (or is replaced by) B when preceded by X and followed by Y'.

16 See Emeneau 1970, Zvelebil 1970 for further details.

17 DEDR = Burrow and Emeneau (1984).

18 It is not the case that we can never reconstruct variation. Evidence for variation can be found in attested ancient languages such as Vedic Sanskrit (e.g. Hock and Pandharipande 1976), as well as in reconstructed languages like Proto-Dravidian (e.g. Zvelebil 1970). However, it is often difficult to recover the social contextual factors which govern such variations.

19 We must not ignore, of course, the role of written symbols in interregional trade even in prehistoric times – such as the use of Harappan seal writing to identify parcels which traveled between different Harappan urban sites and even to more distant locations (see Ratnagar 1981).

20 Though the definition of genetic relationship discussed here is probably accepted nowadays by the majority of historical linguists, it is not the only possible view of the situation. As Colin Renfrew has pointed out, the Russian linguist N. S. Troubetskoy questioned the existence of an original unitary Indo-European language, believing that the resemblances which were thought to prove its existence might have come about through a long period of parallel developments and convergence (see **1.22C, 4.22, 4.25–6**): "There is . . . no powerful ground for the assumption of a unitary Indogerman protolanguage . . . It is just as plausible that the ancestors of the Indogerman language groups were originally quite dissimilar, and that through continuing contact, mutual influence and word borrowing became significantly closer to each other, without however going so far as to become identical" (Troubetskoy 1939, quoted in Renfrew 1987: 108). Though Troubetskoy perhaps represents the minority view, it is difficult to deny the possibility that some cases of accepted proto-languages may be as he has described. Indo-Europeanists might rebut his claim by pointing to the very close similarities in grammatical, phonological, and accentual rules which characterize Sanskrit, Latin, and Greek, the most ancient of the Indo-European languages – similarities which seem detailed enough to allow us to infer that they all derive from a prehistoric speech community which shared "the same norms in regard to language" (see the definition of *speech community* in **1.22**). Even if this is true, however, questions have been repeatedly raised about some branches of the Indo-European family (Germanic and Celtic, in particular) as to whether they are full-fledged members of the family, or were perhaps based on some other linguistic substrata. In all honesty, we must accept that these questions are unresolved – both the specific questions about Germanic and Celtic (and similar cases in other language families, such as Bhili in Indo-Aryan), and the general question about the assumption of uniform proto-languages.

21 Cf. DEDR 333 Tamil **ā(<u>n</u>)** 'be': Brahui **an**; DEDR 3376a Tamil **tūnku** 'hang, swing, sleep': Brahui **tu<u>gh</u>** 'sleep, dream'; DEDR 3690a Tamil **nīr** 'water': Brahui **dīr**;

DEDR4096 Tamil **pāl** 'milk': Brahui **pāl̲h**, as well as grammatical items such as negative verbal suffixes and nominal suffixes (e.g. plural, objective case).

22 That is, if not geographically distinct, the two speech communities had more internal similarity than either group had to the other.

23 When political boundaries are imposed on such regions, it often turns out that the varieties of speech spoken near a boundary are as similar to those on the other side of the boundary as they are to those within it; and if standard languages of education and administration are imposed within each region, those living near the boundaries often find it easier to speak with their neighbors across the border than with visiting officials who speak to them in "their own" language (often based on the speech of a distant capital city). This is a situation found in earlier times in border regions such as that between Holland and Germany, between France and Italy, and across many state borders in South Asia.

24 For a real-life example, modern western Hindi shares several changes with languages to the east, including the merger of retroflex **ṇ** with dental **n**, and the merger of word-initial **v** and **b**, while it shares changes **ai** → **è** (open **e**), **au** → **ò** (open **o**) with languages lying to the west, yielding comparisons like the following: Panjabi **vīs**, Hindi-Bihari **bīs** '20', Panjabi-Hindi **bèl**, Bihari **bail** 'bullock', Panjabi **vènga**, Hindi-Bihari **bèngan** 'eggplant'.

25 A widely used definition of language contact is that of Uriel Weinreich: "…two or more languages will be said to be **IN CONTACT** if they are used alternately by the same person. The language-using individuals are thus the locus of the contact" (1974: 1).

26 In this case, it is only the second element which has been transferred through linguistic interference (Weinreich 1974: 52).

27 A phonological example may illustrate this more clearly. The Romance words for '100', for example French **cent/sã/**, Italian **ciento /čénto/**, Spanish **ciento /siénto/**, Sardinian **kéntu**, etc., and other words with similar sounds, such as the words for 'wax' (French **cire /sir/**, Sardinian **kéra**, etc.) and 'field' (French **champs /śã/**, Spanish **campo**, Sardinian **kámpu**, etc.) lead to reconstructions beginning with **k: *kentu, *kēra, *kampu**, etc. (Hall 1976: 67), which is believed to correspond to the classical Latin pronunciation of these words (Sturtevant 1977). At the same time, the initial **h-** in Latin words like **hostis** 'army', **hospis** 'guest, host', **homo** 'man' is not retained in the corresponding words in the Romance languages: the Proto-Romance reconstruction based on Rum. **oaste,** Old Ital. **oste,** Fr. **hôte** /ot/, Span. **ueste** would be ***oste**, as opposed to Latin **hostis, hoste-**; from Rum. **oaspe,** Ital. **ospite,** Fr. **hôte**, Span. **uesped** 'guest, host' the PRom. reconstruction is ***ospite**, as opposed to Latin **hospis, hospite-;** French **homme/om/**, Spanish **hombre /ómbre/**, Italian **uomo**, etc. yield PRom. ***omo** (Hall 1976). Thus Proto-Romance is reconstructed with ***kentum** '100' and ***omo** 'man', whereas classical Latin had **kentum** (written **centum**) and **homo**, with **h-**. On the other hand, these reconstructions may correctly reflect the reality of the *spoken* language from which the modern Romance languages are descended. As Hall notes: "The Romance languages do not descend in a direct line…from Classical Latin. Proto Romance was… a 'sister-[language]', not a 'daughter-language' of Classical Latin; and the Romance languages are therefore 'nieces' and not 'grand-daughters' [of Latin]" (1976: 177).

This does not, in fact, invalidate the reconstructions, since it is likely that the Latin initial **h-** began to be lost in speech even while it was still retained in writing (as it is still retained in French and Spanish), and this may well have happened in some areas before the various changes of **k** to **č**, **s**, **ś**, etc. Thus there probably existed Rumanic dialects in which descendants of words like **kentum** (with an initial **k**-sound) coexisted with words like **omo** (without initial **h**). All the same, it is probably not the case that there was any period in which this situation existed throughout the whole region in

which Romance languages were spoken; in other words, there was no time when all the regional varieties of pre-Romance had both **kentum** and **omo**. To this extent, then, Proto-Romance must be regarded as a fictional composite language.

28 The Dravidian word for 'horse' is an extreme example of such a case. Burrow (1972) reconstructs a Proto-Dravidian word for horse on the basis of a two-word comparison: Classical Tamil **ivuli** and Brahui **hulli**. These two words, coming from the geographical extremes of the Dravidian languages (southern India and Baluchistan), would seem to be solid evidence for a Proto-Dravidian word, were it not for the shaky phonological resemblance. Krishnamurti (2003: §1.2.2.5) rejects it on the grounds that the phonological correspondences are not the expected ones. Therefore, in the interests of scientific accuracy, the word must be listed with a question mark, and no inferences can be drawn from it.

29 The distribution of these variants is more complex than implied by the opposition of "working class" vs "middle class". This is an example of what William Labov and others have called **negative concord**, a stable sociolinguistic variable which shows the type of distribution illustrated in Figure 1.3. See Labov (1966, 2001: 78ff.).

30 "…/in/ is the direct and regular descendant of the Old English participial ending **-inde**, **-ende**, just as/iŋ/ is the regular result of sound changes operating on the Old English verbal noun **-inge, -ynge**." (Labov 2001: 88; see also Houston 1985, 1991). Note the use of the participle in **-ende** in this quotation from Old English; compare German verbal nouns like **Ordnung, Rechnung, Űbersetzung**, and participles like **gehend, laufend, űbersetzend**.

31 Strictly speaking, regional variation is also probabilistic in some circumstances, especially in transitional areas between dialects.

32 Cross-cultural studies of language and gender show the situation to be complex. A number of languages of pre-industrial societies show marked differences between male and female speech. In Yana, a California language (Sapir 1949), different forms of many words are prescribed for men (when talking to men) and for speech by and to women; the examples given suggest that the differences may have originated in the reduction or elimination of final syllables by female speakers. For Koasati (southeastern US), on the other hand, Haas (1944) found that it was descriptively simpler to posit female forms as basic and to derive the male forms from them (though this does not necessarily reflect historical reality). In Kudux, the speech of a patriarchal 'tribal' group in eastern India, some of the forms used among women seem to reflect an older stage of the language (Ekka 1972). See Bonvilain 2000: 209–34 for discussion of these and other cases.

33 One notable exception was a group of married women in one village who commuted daily by bus to a nearby town where they worked as prostitutes.

34 The Tamil words for 'rice' illustrate this clearly: the word **arici** 'rice in the field, raw rice' is shared by all Tamil speakers, whereas the words for 'cooked rice' vary in terms of caste membership: Brahmins and other high-caste members mostly use the term **cātam**, while members of lower castes say **cōru**.

35 It is possible that many of us unknowingly use each others' terminology without being fully aware of the denotations and connotations they have for scholars in the other field. I myself have been guilty of using terms like 'archaeological culture' without knowing exactly how an archaeologist would define them.

2

THE SOUTH ASIAN LINGUISTIC SCENE

Contents

2.1. Introduction

This chapter presents background information on the linguistic situation in South Asia, starting from the earliest period for which useable data are available. As there are detailed treatments of this subject elsewhere, no attempt is made at completeness here.[1] Three South Asian language families provide data for the prehistoric period: (1) *Indo-Iranian*, with three branches: (A) Indo-Aryan, with texts in Old Indo-Aryan or OIA dating back to the middle of the second millennium BCE,

(B) Iranian, with texts in Avestan dateable to about the same period, (C) Nuristani (or Kafiri), a group of non-literary languages spoken by small groups in the mountain ranges of the northwestern part of the subcontinent; (2) *Dravidian*, for which comparative evidence provides reconstructions dating back probably to the third millennium BCE; and (3) *Munda*, a branch of the Austro-Asiatic family, which provides reconstructions dateable to about the middle of the second millennium BCE. Each of these major linguistic groups is discussed here briefly.

An important linguistic aspect of the region is that most South Asian languages, regardless of their genetic affiliations, share a number of structural features, and thus one may appropriately speak of South Asia as a "linguistic area" (Masica 1976).[2] These features include extremely detailed resemblances in word order such as SOV (Subject–Verb–Object) order in unemphatic declarative sentences, postpositions, left-branching Noun Phrases, and syntactic processes like the "dative-subject" construction (as in Marathi **ma-lā t-o dis-l-ā** ["me-to that-Masc. appear-Past-Concord"] = 'I saw it/him'), **t-i ma-lā āvaḍ-l-i** ("that-Fem. me-to please-Past-Concord") 'I liked it/her', the extensive use of "conjunctive participles" for conjoining sentences (as in Hindi **jā-ke dekh**, Tamil **po-i pāru** ["go-Ppl look"] = 'go and look'), and "double causative" verbs (causative of causative, e.g. Hindi **banvā** 'cause to cause to be made', from **banā** 'make, cause to be made', from **ban** 'be made, become'). Masica's study of areal features shows that a number of these "Indian" features are also found widely distributed in a large Eurasian area to the north of South Asia. On the other hand, the areas to the immediate east and west of South Asia – Southwest Asia (the "Middle East") and Southeast Asia, respectively – show much less linguistic similarity with South Asia. (See Masica 1976: 178–84 and map 7, p. 181.)

2.2. Indo-Aryan

Indo-Aryan languages are spoken by upwards of 650,000,000[3] people (as of 1981), spread over a contiguous area which covers most of Pakistan, the northern two-thirds of India, and Bangladesh (see Figure 2.1).[4] Within this area there are various groups – some of moderate size, most extremely small – speaking languages of three other families: Dravidian, Austro-Asiatic and Sino-Tibetan, as well as a few isolated languages (see **3.2**). The major literary Dravidian languages are found in the southern part of India, also known as "peninsular India" (see Figure 2.3).

2.21. Old Indo-Aryan (OIA)

The oldest[5] specimens of an Indo-Aryan language in South Asia are the hymns of the Rigveda (OIA **ṛgveda**), a compilation of sacred verses (OIA **ṛc**) which on internal linguistic evidence were composed at different epochs. These hymns are the most sacred scriptures of Hinduism, still used today for marriages, funeral rites, death ceremonies, and other ritual occasions. On the basis of the names of rivers mentioned in the hymns, it is presumed that the geographical focus of the Rigvedic society was the **sapta-sindhu**, the region of the seven northern tributaries of the Indus,

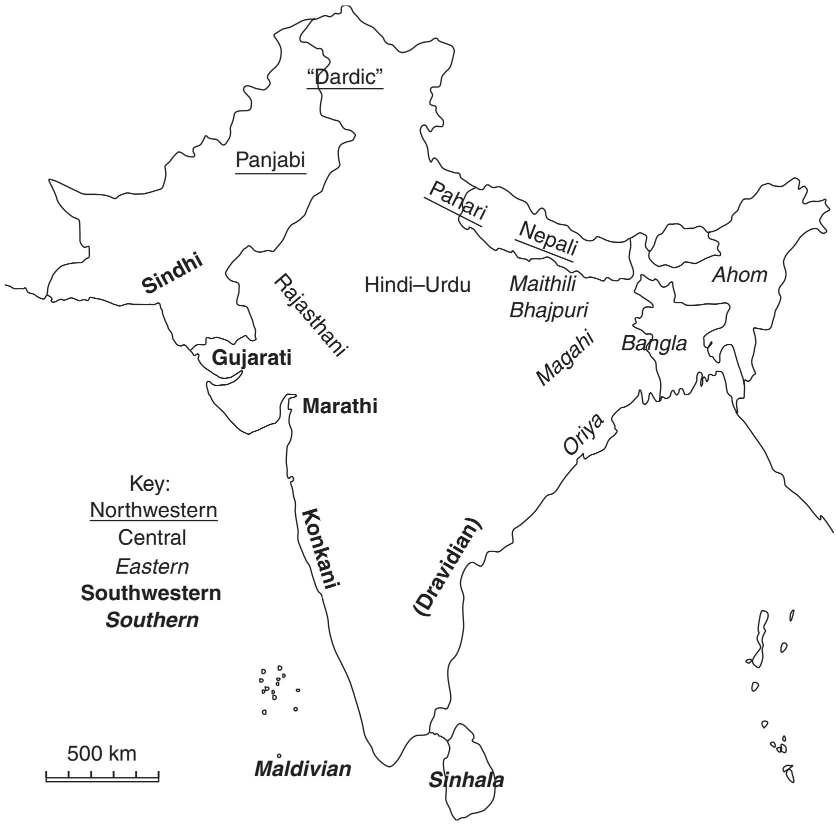


Figure 2.1 Major Indo-Aryan languages.

"... roughly from the Kabul river to the Sarasvati river", now the dry bed of the Ghaggar (Thapar 1978: 214). Since the Rigveda shows no evidence of any substantial contact with the Harappan civilization of the Indus Valley, it would follow that the Vedic people entered the Indus Valley only after the decline of that civilization during the early second millennium BCE. The references to battles and destruction of fortresses, once thought to represent the destruction of the Harappan cities by Aryan hordes (Wheeler 1968: 131–2), if they have any basis in fact, are more likely to refer to small-scale attacks on walled villages of post-Harappan peoples (see Thapar 1978: 215, 1984: 43–4), or possibly even to central Asian forts (see Erdosy 1995b: 11–12).

An approximate confirmation of the mid-second millennium date for the composition of the Rigvedic hymns is provided by fragments of the Old Indo-Aryan language found in an inscriptional treaty of the Mitanni of the upper Euphrates, dateable to about 1380 BCE. The treaty includes names of Mitanni princes which are of Indo-Aryan derivation, along with the names of the gods Mitra and Varuna, Indra and Nasatya, clearly identical with the OIA **mitra-varuṇa** (regularly paired

in the Rigveda), and **indra-nāsatyā** (found in one Rigvedic hymn). In addition, some Indo-Aryan terms for horse-racing are found in a Hittite manual on horse-breeding of the same period (Bloch 1965: 1; Mayrhofer 1974). These facts are generally interpreted to imply that a group of OIA speakers, roughly contemporary with the groups which created the Rigveda, occupied an elite niche in the Hurrian-speaking society of that time. Since nothing further is heard from them, it is presumed that they were ultimately absorbed into the local population. Since the language of the earliest Rigveda appears to be slightly more archaic than that of the Mesopotamian fragments, it may be placed a century or so earlier.

The Rigvedic hymns depict a pre-urban society with a copper technology,[6] evolving from nomadic pastoralism dependent on cattle to a form of settled agriculture.[7] Tribal or lineage identity was important, and the patriarchal family was the basic social unit (Thapar 1978: 214). There are indications of "cultural exclusiveness and separation from the local people" (ibid.), and expressions of contempt for the physical appearance, language, and religious beliefs of the indigenous people, the Dasas (OIA **dāsa**) (see **2.84**, also Deshpande 1979b: 2–3).

2.21A. Prehistory of OIA

As every student of historical linguistics learns, it was the discovery of Sanskrit by Europeans, and its comparison with the classical languages of Europe, which led to the development of the field of comparative-historical linguistics in the nineteenth century. As more and more detailed resemblances among the ancient languages Greek, Latin, and Sanskrit were revealed, scholars became convinced of the truth of William Jones's claim of 1786 that all three languages "have sprung from some common source which, perhaps, no longer exists" (see Pedersen 1931). For modern-day skeptics, it should suffice to look at equivalences like the following to recognize that there must be a fundamental, non-accidental historical connection among them:

Sanskrit	*Latin*	*Greek*	(*meaning*)
ásti	est	estí	'(he/she/it) is'
sthá	estis	esté	'(you) are'
sánti	sunt	entí	'(they) are'
pitar	pater	patér	'father'
daśa	decem	déka	'ten'
bhar-āmi	fer-ō	fér-ō	'I carry'

In any case, there is no doubt among historical linguists that these three languages – along with those of the Balto-Slavic, Celtic, Germanic, and other groups – are genetically related in the sense of **1.22A**. At the same time, great caution must be used in drawing historical conclusions from such linguistic relationships. It is regrettable that from the time these relationships were discovered, scholars and others have used them as a basis for very far-reaching conclusions about the prehistoric sources of cultural, ideological, and racial characteristics of people. Linguistic archaeology uses

the evidence of genetic relationships to infer the existence of prehistoric speech communities – not races, ethnic groups, or cultures (see **1.22B**). In the case of Proto-Indo-European (PIE), we infer that the language of this speech community was the predecessor of all the Indo-European (or Indo-Hittite) languages.[8]

Before proceeding further, a disclaimer of sorts is needed. As outlined in section **1.2**, the demonstration of a genetic relationship between languages – which, as noted there, has nothing to do with human biology – implies the existence of a prehistoric speech community which was the predecessor or "ancestor" of the languages shown to be related. Given similarities of the kind illustrated here, no other assumption is possible. However, though we assume the existence of prehistoric speech communities speaking proto-Indo-European, proto-Indo-Iranian, proto-Indo-Aryan, etc., it is extraordinarily difficult to demonstrate their existence by material evidence, and equally difficult to find the links between the widely dispersed existing speech communities which would tell us how they came to be where we find them. The assumption that speakers of Indo-Aryan languages entered the South Asian subcontinent from Central Asia in the second millennium BCE has been questioned by archaeologists who claim to find no evidence of intrusive elements which could be identified with speakers of these languages – or in fact any languages. This problem is not unique to the South Asian case. The author's belief is that the first stage in resolving this question is to turn our attention to the process of *migration*. We do not know how languages migrated in the distant past. We do not know enough either about the numbers of people required, or about the social processes involved, nor do we know what signs of movement we should look for. This is a subject which will only be resolved by joint work between archaeologists and historical linguists. (See further discussion in Section **10.3**.)

Having said this, it will be assumed in the intervening chapters that the Indo-Aryan languages entered the South Asian subcontinent from the outside, even though we cannot specify the process by which this happened.[9] On this assumption, the early history of the Indo-Aryan languages can be briefly summarized as follows: Indo-Aryan and Iranian constitute the Indo-Iranian branch of the Indo-European languages. The Proto-Indo-Iranian speech community is believed to have split into a southern (Proto-Indo-Aryan) and northern (Proto-Iranian) branch about 2000 BCE, in a region somewhere between Central Asia, that is, ancient Bactria–Margiana or roughly modern Tadjhikstan–Uzbekistan, and the area south of the Ural Mountains. The speakers of Proto-Indo-Aryan subsequently moved through Central Asia into what is now Iran and Afghanistan, later to the upper Indus Valley, and later spread to the northern part of what is now India. (See **10.3** for further discussion; see Burrow 1973a for a detailed account, and Masica 1991: 34–7 for a summary.)

2.21B. Sources for OIA

In addition to the Rigveda, there are three other collections of material known as the Atharvaveda (OIA **atharvaveda**), the Samaveda (OIA **sāmaveda**), and the Yajurveda (OIA **yajurveda**) which exists in two recensions. The Atharvaveda, the

Veda (= knowledge, cf. English **wit**) of the **atharvans** or fire-priests, is basically a compilation of magical texts. While the language of the Atharvaveda contains some forms which are later than those of the Rigveda, it is generally believed that much of the language, as well as the content, of the Atharvaveda is as old as that of the Rigveda (see Whitney 1950: xvi; Emeneau in Dil 1980: 180). In fact the Atharvaveda appears to represent a separate, and perhaps competing, tradition – a "magical" or possibly shamanistic tradition, which was only later integrated into the mainstream of Hindu religious practice. The remaining Vedas are of a liturgical nature: the Samaveda is a collection of metrical songs or chants, mostly made up of verses from the Rigveda, while the Yajurveda is a collection of Rigvedic mantras, or ritual formulas, for use by the sacrificial priests.

In addition to the four Vedas, the literature of the Vedic period includes prose materials known as Brahmanas (OIA **brāḥmaṇa**), containing various materials relating to the Vedic sacrifices, along with myths, philosophical speculations and the beginnings of scientific thought in the areas of astronomy, grammar, and phonetics. Somewhat later are the Aranyakas (OIA **āraṇyaka**) or wilderness books, and the Upanishads (OIA **upaniṣad**) or lectures on sacred subjects. The four Vedas along with the Brahmanas, Aranyakas and Upanishads are the main constituents of what is known generally as the literature of the Vedic period, though the Upanishads were continued and added to down to relatively modern times; they are according to Whitney "one of the lines by which the Brahmana literature passes over into the later theological literature" (1950: xix).

By the end of the Vedic period the geographical focus had shifted to the middle Ganges or Ganga (OIA **gangā**), and the lineage system of the Rigveda had shifted to a "combined lineage and householding economy" (Thapar 1984: 17), an important step in the transition toward a state economy. Because of the great development in culture, religion and language between the Rigveda and the Upanishads, it must be assumed that a number of centuries elapsed between the end of the Rigvedic period and late Vedic times. Basham (1954: 31) suggests 500 years, a reasonable and convenient figure, which would bring the end of the Vedic period as a whole to 700 BCE at the earliest, even assuming with Deshpande (1979b: 1) that the Rigvedic period ended about 1200 BCE. The date of 700 BCE is not far from the beginnings of the teachings of the heterodox sects – Buddhist, Jain, and others, beginning with the Buddha (fifth to fourth century BCE) – which "presuppose the entire Veda as a literature to all intents and purposes completed" (Winternitz 1959: 288).

The term "Classical Sanskrit" (or often "Sanskrit" alone) refers generally to the post-Vedic form of OIA, and more particularly to those works which follow the rules of the grammar of Panini (**pāṇini**), dateable to the fifth century BCE (Whitney 1950: xiv; Winternitz 1959: 38–9, note Winternitz's term "ancient Sanskrit" to refer to Vedic prose passages). Lexically and grammatically, the later language differs considerably from that of the Vedic hymns; simplifications and other differences in the verbal system are especially noticeable (see Whitney 1950: 200ff.; Bloch 1965: 207ff.). Classical Sanskrit flowered in the late centuries BCE and in the early centuries of the CE.

Epic Sanskrit, the language of the great Hindu sagas, the Mahabharata (OIA **mahābhārata**), the story of the great war of the Bharatas, and the Ramayana (OIA **rāmāyaṇa** 'travels of Ram'), can be to some extent included in classical Sanskrit. Much of the language of these two epics, however, fails to conform to the Paninian standard, often showing the influence of Middle Indo-Aryan. In addition, there are archaisms in the epic language which suggest earlier origins. Hock and Pandharipande note that Panini makes reference to a **(mahā)bhārata**, indicating that some form of this work may have already existed in his time, though the available versions probably represent a compilation dating to the early centuries CE. They suggest that the epic tradition may have established its own norms "before Panini and his normalizing influence" (1976: 120). This is highly likely, given that the Mahabharata stories are based on a bardic tradition which appears to go back to Vedic times (Winternitz 1959: 436ff.).

2.22. Middle Indo-Aryan (MIA)

MIA includes Pali (**pālī** or **pāḷī**), the language of the Buddhist canon and other major Buddhist writings, as well as the dialects of the Ashokan inscriptions of the mid-third century BCE and other inscriptional Prakrits, the dramatic Prakrits which are used to represent the speech of most characters in the classical Sanskrit dramas (while the high-class male protagonists spoke Sanskrit), and the literary Prakrits which were used for both prose and poetry. The term Prakrit (OIA **prākṛta** 'original, natural, normal, artless') is opposed to **saṃskṛta** 'refined, elaborated'. The Prakrits are generally regionally colored varieties which can be to some extent localized geographically and chronologically, but whose relationships with actual spoken varieties are vague at best. (See further discussion in **2.8**.)

The varieties known as Middle Indo-Aryan are, from the point of view of OIA, considerably evolved both phonologically and morphologically. The most prominent phonological changes include (1) the merger of the vocalic **ṛ** with one of the vowels **i, a, u** in various phonological environments in different MIA varieties: thus OIA **ṛkṣa** → MIA **accha**, OIA **ṛṣi** 'sage' → MIA **isi**, OIA **ṛṣabha** 'bull' → MIA **usabha**; (2) the assimilation of consonant clusters into geminates or doubled consonants non-initially in words, and their reduction to single consonants word-initially: for example, OIA **kartavya** 'to be done' → MIA **kattavva-/kaṭṭavva**, MIA **satta** ← OIA **sapta** '7', **šakta** 'powerful', **šapta** 'accursed'; OIA **traya** '3' → MIA **tayo**; (3) merger of the three OIA sibilants **š ṣ s** → MIA **s** (see the previous example and also OIA **ṣodaša** '16' → MIA **soḷasa**), and (4) loss of all final consonants.[10] Morphological differences include, in noun morphology, the loss of the dual number and reduction in the number of declensions; in the verb, gradual reduction of the number of verbal conjugations and loss of irregular forms, along with substantial loss of verbal distinctions: the middle voice is mostly lost, and apart from scattered remnants the perfects, imperfects and aorists disappear.

A later group of literary languages, the Apabhramshas (OIA **apabhraṁśa** 'breaking down, corruption', originally applied to incorrect or non-Aryan forms

of speech), represent a still later linguistic stage of MIA, phonologically and morphologically further eroded than the Prakrits.

2.23. Modern Indo-Aryan (NIA)

Writings in early forms of the modern Indo-Aryan languages such as Marathi (**marāṭhi**) and Bengali or Bangla (**bāṅglā**) begin to appear around the thirteenth century CE (see Bloch 1919: 35; Chatterji 1970: 109). However, inscriptions in Marathi are known from the tenth century (Tulpule 1960: 84), and Chatterji (1970) relates place names and other materials from as early as the fifth century to modern Bangla. Geiger (1938: 6) dates modern Sinhalese or Sinhala from the middle of the thirteenth century, but comments that the "proto-Sinhalese" of the fifth to eighth centuries [shows] "all the important linguistic rules . . . which are dominant in the modern language and decisive for its character" (1938: 3–4).

The regional groupings of Indo-Aryan languages shown in Figure 2.1 are convenient for reference purposes, but do not represent genuine linguistic subgroups of Indo-Aryan. See Chapters 5 and 6 for discussion of this question.

2.24. Chronology of Indo-Aryan

Figure 2.2 presents a rough chronology of the Indo-Aryan languages. For the Vedic period, the dates represent the probable times of compilation of the various works, since these works were preserved by oral tradition for many centuries, and only committed to writing at much later dates. (While as suggested here, the Vedic period as a whole probably fits approximately into the period between 1700 and 700 BCE, the placing of individual works within this period is largely a matter of guesswork.)[11]

2.3. Iranian

Indo-Aryan and Iranian languages together form the Indo-Iranian branch of the Indo-European languages. Avestan, the language of the sacred writings of the Zoroastrians, contains in its oldest texts, the Gathas (OIr **gāthā**), a language which is generally regarded as contemporary with that of the Rigveda. A later form, known as Old Persian, is preserved in the inscriptions of Darius and other Achaemenid kings. Avestan is in fact so close to Vedic, not only linguistically but also in terms of the cultural and religious concepts it reflects, that some verses in the Avesta can be converted into Vedic verses by fairly minor phonetic substitutions (Burrow 1973a: 3ff.).

2.31. Kafiri/Nuristani

The Nuristani or Kafiri languages, a handful of minor languages found in Eastern Afghanistan and in the Chitral region of northwestern Pakistan, form a third branch of Indo-Iranian. These communities have preserved both linguistic and cultural features which belong to the Proto-Indo-Iranian stage (pre-Hindu and pre-Zoroastrian). (See Morgenstierne 1973a; Witzel 1999b; Degener 2002.)

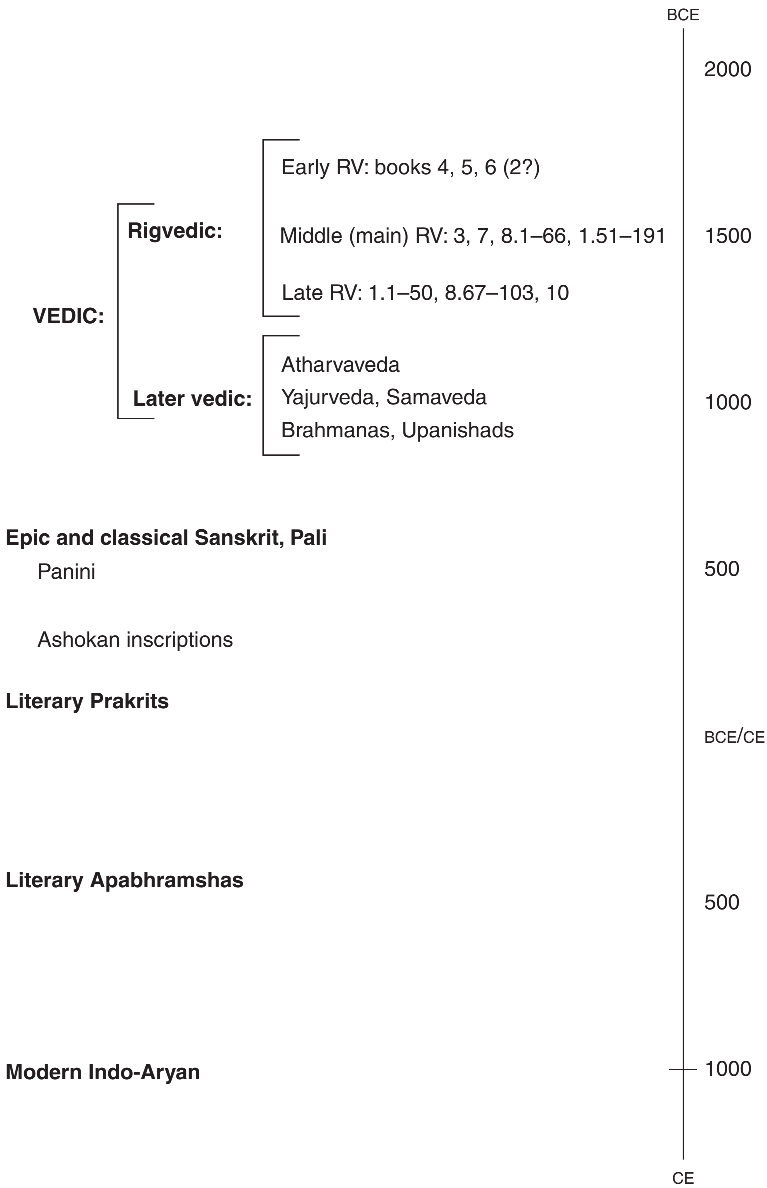


Figure 2.2 Chronology of Indo-Aryan languages.

2.4. Dravidian

The Dravidian language family, comprising some 25 or so languages spoken mainly in South and Central India (see Figure 2.3), includes the literary languages Tamil, Telugu, Kannada, and Malayalam, some of which have records going back to the early centuries CE, as well as a number of non-literary languages spoken by groups which were still living under pre-industrial conditions until very recently. The *Dravidian Etymological Dictionary* (Burrow and Emeneau 1984; abbreviated here as DEDR) contains all the known words which show relations between any two or more Dravidian languages, assembled under approximately 5,500 entries.[12]

Dravidian languages (see Figure 2.3) were for a long time considered indigenous to South Asia, and it is possible that they are, though several recent studies have raised the possibility that they were spoken earlier in areas to the west or



Figure 2.3 Approximate locations of modern Dravidian languages.

northwest of the subcontinent (see e.g. Tyler 1968; McAlpin 1981). McAlpin's proposal of a relationship between Dravidian and Elamite, an ancient language of western Iran, would imply the possibility that speakers of Dravidian languages entered the subcontinent as late as the third millennium BCE. (See further discussion in Chapter 8.) Tyler's (1968) paper on Dravidian and Uralic suggests a possible northern origin of Dravidian. It is not beyond the bounds of possibility that Dravidian is related to *both* Elamite and Uralic; we do not know enough about either to make such a determination. See further discussion in **8.41**.

Dravidian languages are first mentioned in the Sanskrit epics around the middle of the first millennium BCE. Later records show various Dravidian-speaking groups in South India in the late centuries BCE and early centuries CE. Investigation of earlier locations of Dravidian speakers depends heavily on careful sifting of etymological evidence (see **3.22**); agricultural and botanical terms are an important part of that evidence (see **7.21**, **8.4**).

There are three periods and areas in question:

(1) Dravidian languages are now found mainly in peninsular India, in the four southern states of Andhra Pradesh, Karnataka, Kerala, and Tamilnadu. As argued in **8.4**, a good case can be made for an association between the Proto-Dravidian speech community and the Southern Neolithic archaeological complex of the mid-third millennium BCE, which would imply the dominance of Dravidian languages in the peninsula from that period onward.

(2) There is linguistic and cultural evidence for the presence of Dravidian speakers in western India, in the states of Maharashtra, Gujarat, and Sindh, at earlier periods (see **9.4**, **9.5**). In addition, there are loanwords in Old Indo-Aryan suggesting contact with Dravidian speakers (possibly in Sindh and southern Panjab) starting from the middle Rigvedic period (*c.*1200 BCE); Section **3.22A** contains some previously unpublished additions to these data, which suggest that this contact may have started somewhat earlier. These Dravidian-speaking populations may have been linked to those of the Southern Neolithic complex mentioned in (1), or may have been the descendants of earlier arrivals (see next paragraph) – or possibly both.[13]

(3) The possibility that Dravidian languages are related to Elamite and/or Uralic has not been proven, but cannot be discounted. Reinforcing that possibility is the (still very controversial) linguistic evidence for contact between Dravidian speakers and the proto-Indo-Iranian speech community, including some new evidence (see **3.22A**, **3.23**, **3.32**, and Chapter 3 Appendix).

2.41. Subgroupings of Dravidian

Figure 2.4 shows two possible subgroupings of the Dravidian languages. The left side of the figure shows the generally accepted classification, with three primary subgroups: North Dravidian (ND), Central Dravidian (CD), and South Dravidian (SD); South Dravidian is further subdivided into two subgroups known as South

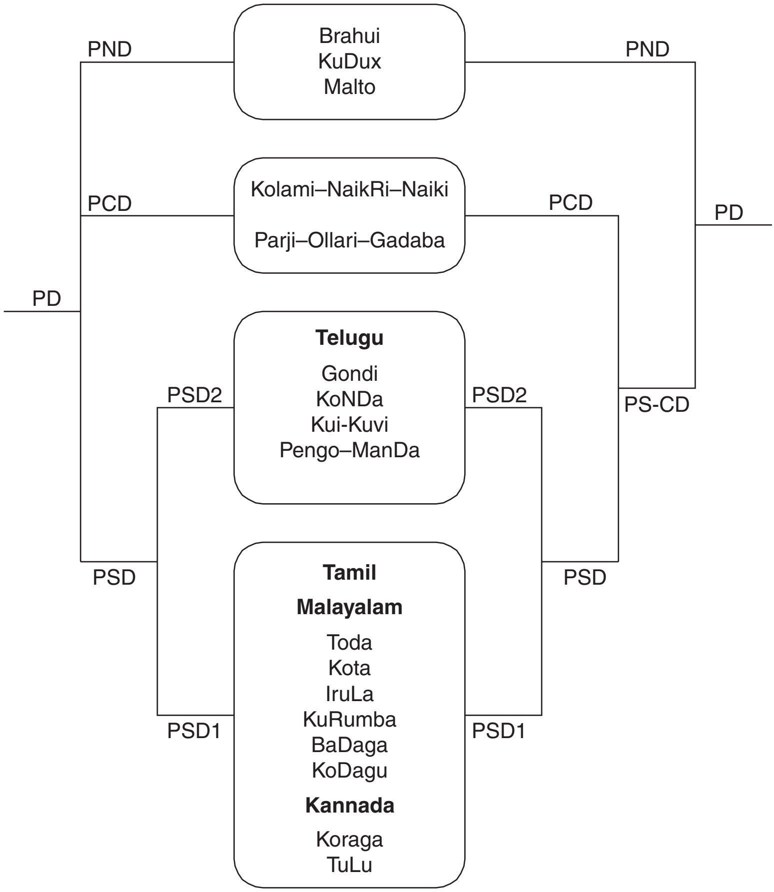


Figure 2.4 Alternative subgroupings of Dravidian languages (see text for discussion).

Dravidian-1 (SD1) and South Dravidian-2 (SD2). The right side of the figure shows an alternative classification which posits two major subgroups: North Dravidian and South/Central Dravidian, with the latter further subdivided into Central Dravidian and South Dravidian – which further subdivides, as in the first scheme, into SD1 and SD2. (This figure is isomorphic with figures 11.2A and 11.2B in Krishnamurti 2003.) This matter is dealt with in **8.22**.[14]

2.42. Dating

Dating reconstructed languages is notoriously precarious, and the most that can be expected in such cases is a vague approximation, or perhaps a plausible range of dates. On the basis of changes attested in the earliest written documents, a date

in the second half of the first millennium BCE can be suggested for PSD1 (see Krishnamurti 1969: 317). Krishnamurti (2003) cites early references in Sanskrit, one in the Aitareya Brahmana (*c.* seventh century BCE) which mentions the tribe **Andhra**, presumed to be Dravidian-speaking, and another in Bharata's Natyashastra (*c.* fourth century BCE) which mentions the languages **Dramila** (= Tamil?) and **Āndhra** (= Telugu?), and concludes as follows:

> We can, therefore, infer that the split of South Dravidian I (with pre-Tamil as the dominant language and South Dravidian II (with pre-Telugu as the dominant language) could precede the period of the Aitareya brāhmaṇa by at least four or five centuries, i.e. ca. 10–11th century BCE.
>
> (Krishnamurti 2003 §11.4, reprinted by permission from Bh. Krishnamurti, *The Dravidian Languages: A Comparative, Historical and Typological Study*, Cambridge: Cambridge University Press)

If this argument is accepted, then we would be justified in placing the first breakup of Proto-Dravidian as far back as 1500 BCE, or perhaps as far as 2000 BCE, assuming a tripartite division (lefthand side of Figure 2.4). This is only the most crude estimate of time, based on vague notions of how long major linguistic innovations require for their completion – and recognizing that the duration of such changes is highly variable.[15] On the other hand, the occurrence of probable Dravidian loanwords in OIA texts of the middle Rigvedic period, starting around 1200 BCE (see **3.22** and Witzel 1999b), suggests that the date of 2000 BCE for Proto-Dravidian may not be too wide of the mark. In fact, it is possible that Proto-Dravidian is older than this (see **8.4**).[16]

2.5. Munda and Austro-Asiatic

Munda languages (Figure 2.5) are at present located mainly in eastern India,[17] and there is no direct evidence for their having been more widespread in earlier times, though the location of Korku/Kurku on the middle Narbada could possibly point in that direction, and the evidence of loanwords in Vedic and later Indo-Aryan points to the probability that speakers of Munda and/or other related languages were spoken in the prehistoric Panjab and the Ganga–Yamuna *doab* (Witzel 1999b; see **3.21**). The Austro-Asiatic family, of which Munda is a branch, includes various languages of Southeast Asia, as well as Khasi (spoken in Assam) and Nicobarese (in the Nicobar Islands in the Bay of Bengal).

Blust, accepting the 'Austric hypothesis' which posits a genetic relationship between Austronesian (AN) languages and Austro-Asiatic (AA) languages, places the early habitat of the larger Austric family "in the general region where the Salween, Mekong and Yangzi rivers run parallel on the Burma-Yunnan frontier" (1996a: 136). He suggests that by 7000–7500 BP, PAA (Proto-Austro-Asiatic) had separated into western (Munda) and eastern (Mon-Khmer) dialect areas

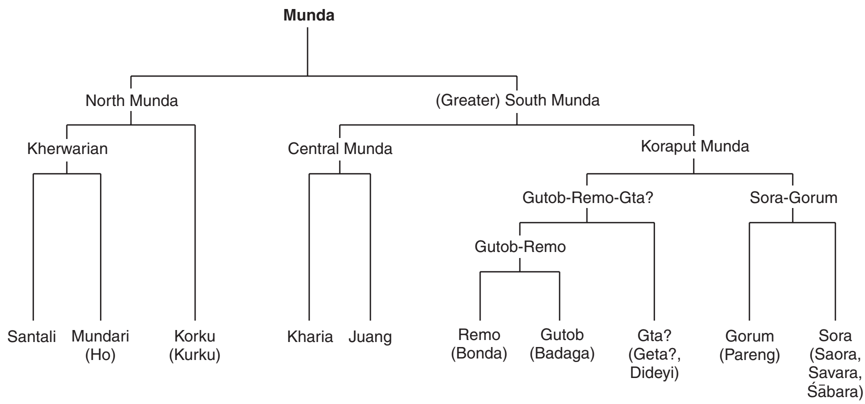


Figure 2.5 Subgrouping of Munda languages (information provided by David Stampe, 2003).

(1996a:132); "Munda [moved] down the Brahmaputra Valley into Assam and Bangladesh, gradually spreading westward, and Mon-Khmer down the Salween and Mekong valleys into mainland Southeast Asia." The subsequent spread of Tibeto-Burman-speaking peoples down the Irawaddy and Salween Valleys would have separated Khasi from Mon-Khmer (1996a: 137).

2.6. Tibeto-Burman (Sino-Tibetan)[18]

Tibeto-Burman languages are at present scattered in small pockets in the Himalayan and sub-Himalayan regions. Only a few tentative linguistic links have been established between the languages of this family and the others discussed here. See **3.2(1c)** for further discussion.

2.7. Other languages

Apart from the four families mentioned in **2.2–2.6**, there are a few languages in the subcontinent which show no identifiable relationship with any languages inside or outside of the subcontinent, in spite of some fairly intensive efforts on the part of individual scholars. The best known of these are: Burushaski, a language of Kashmir (Lorimer 1935–38), Nahali, a language of central India (Kuiper 1962), and Vedda, a language of Srilanka (De Silva 1965). Though limited in extent nowadays, some of these languages may have been of greater importance in prehistoric times. In addition, studies of loanwords in OIA and early Dravidian give hints of the existence of other pre-Aryan and pre-Dravidian languages. See **3.2** for further discussion of these languages and their possible implications for the linguistic prehistory of South Asia.

2.8. The sociolinguistic situation in ancient India

2.81. Introduction

While there is easily enough material available on this subject for at least one doctoral dissertation, my purpose here is to discuss briefly two topics: sociolinguistic variation in ancient Indo-Aryan (**2.82–3**), and the evidence for the presence of speakers of non-Indo-Aryan languages (**2.84**), as necessary background for discussions in later chapters.

For the prehistoric/protohistoric period,[19] we have to depend for our knowledge of intergroup communication on evidence available in early Indo-Aryan languages: Vedic, Sanskrit, Pali, and the Prakrits. Much of this evidence is indirect. Apart from words in ancient texts which are identifiable as foreign (see **3.1**), relevant types of evidence include (1) statements by grammarians, (2) references in texts such as the Vedic hymns, (3) the Sanskrit dramas, in which not only Sanskrit but various types of Prakrits appear, (4) variant forms in the ancient texts, (5) evidence of structural influence of other languages on Indo-Aryan (**3.3**).

2.82. Linguistic variation in Old Indo-Aryan

The chronological sequence given in Figure 2.2 for Indo-Aryan refers to the historical sequence in which the various literary forms of Indo-Aryan make their appearance.[20] Both internal evidence and linguistic form are used to place Vedic and Sanskrit chronologically. In addition, certain non-reversible linguistic evolutionary developments determine the chronological relationships: later Sanskrit, for example, differs from Vedic in having reduced and regularized much of its morphology; the Prakrits differ from Sanskrit in the loss of certain phonological distinctions (see **2.22**).

However, it is clear that varieties of speech with "prakritic" phonology or morphology (see **2.22**) existed from the Rigvedic period onward. Hock and Pandharipande cite for example the prakritism **kim** 'what' for older **kad**, both of which appear in RV (1976: 112–13; see further examples and references therein). Furthermore, a number of cases have been shown where the Prakrits retain forms which are *more* archaic than the equivalents in the earliest Vedic, showing that these dialects were continuous from the pre-Vedic period. Thus the Prakrit languages, or dialects, known from early inscriptions (such as those of King Ashoka in the mid-third century BCE), from the early writings of Buddhists and Jainas, and from the early Sanskrit dramas, were contemporary with the Sanskrit of that period. For the earlier Vedic period, though there is no direct evidence for any form of speech other than Vedic, the coexistence of contemporaneous "prakritic" varieties must be inferred. Cf. Emeneau's comment:

> We have an inkling . . . of the oldest Indo-Aryan of North India as a large dialect area whose speakers were unified by a common culture and by

> the religion that provides us with the evidential documents; there probably were other dialects as well, outside of this social and religious milieu...
>
> (1966: 127; reprinted by permission from Murray B. Emeneau, 'The dialects of Old Indo-Aryan', in H. Birnbaun, J. Puhvel (eds), *Ancient Indo-European Dialects*, Berkeley and Los Angeles: University of California, pp. 123–38)

Regarding the question of a Prakrit contemporary with the Rigveda, Emeneau notes:

> Tedesco...prefers to call this dialect parallel to Rgvedic 'archaic Middle Indic'; probably 'Proto-Middle-Indo-Aryan (or Indic)' is as good. No absolute chronology is possible for it...But we can guess that the latter [Proto-Middle Indo-Aryan] type of dialect was in existence from the beginning of the time span.
>
> (1966: 131)

2.83. Diglossia

That this was a diglossic relationship, and not simply dialectal variation, is shown by the existence from early times of "hyper-Sanskritized" forms, that is, Vedic or Sanskrit forms which were treated as if they were Prakrit, and "sanskritized" to avoid the appearance that the speaker/writer was using a Prakrit form. For example, on the basis of the equation Pkt **tt** = Skt **kt**, as in cases like Skt **mukti** 'release': Pkt **mutti**, **šakta** 'powerful' : Pkt **satta** (see **2.2**), a Skt form **muktā** 'pearl' was created from Pkt **muttā**.[21] Such changes indicate the speakers' awareness of the phonological relationships between the high (Vedic or Sanskrit) forms and the low (Prakrit) forms, supporting the notion of the existence of a diglossic relationship between the two forms of speech.

For the period of Classical Sanskrit there is also the evidence of the Sanskrit dramas, in which the upper-class male protagonists spoke Sanskrit, while most other characters spoke various Prakrits. Hock and Pandharipande argue that these different varieties were all expected to be mutually intelligible, and that at least in the early dramas of Bhasha and Ashvaghosha (*c*. first century CE) the Prakrits used were "more archaic and presumably more natural" than the later stylized Prakrits (1976: 114).[22] In fact, the dramas do not depict diglossia as such, since the individual characters stick to their prescribed variety regardless of the circumstances. Thus this juxtaposition of languages may be a mere literary convention rather than any attempt to represent sociolinguistic reality. Lee notes that in the earliest dramas the verses are in Sanskrit and only the dialogues in Prakrit, suggesting that the Prakrit passages had been added "to what was essentially a Sanskrit drama [possibly derived from epic recitations] in order to reflect the status of the inferior characters" (Lee 1986: 153). It is likely that by this period (late centuries BCE

and early centuries CE), Sanskrit was largely restricted to ritual and scholarly use. If so, the Sanskrit portions of these dramas were perhaps related to the spoken usage of their time somewhat as the dramas of Corneille and Racine are related to contemporary French, or those of Shakespeare to contemporary English, or the classicalized language of Tamil films to modern colloquial Tamil.[23]

In the Rigvedic period, at least the earliest part of it, the language of ordinary daily speech can be presumed to have been quite close to that of the texts which have been handed down to us. As with most ritual languages, in a very few generations it would begin to sound archaic and different from ordinary speech, especially as in the case of the Vedas great care was taken to preserve the exact linguistic forms. The prakritic varieties of speech which existed at this time were presumably used by those members of the society who had no opportunity (or were even forbidden) to learn the ritual language – women, children, members of other castes and other groups – and on occasion by the adult Brahman males, especially when interacting with others.

In the later Vedic period the language of the Rigveda was already archaic, and presumably sharply differentiated from the ordinary usage of the time. Whitney characterizes the prose of the Brahmanas as "in most respects intermediate between the classical [Sanskrit] and that of the Vedas" (1950: xviii). Apart from Vedic and Sanskrit, the Prakrits had been diverging for over half a millennium and were presumably quite distinct forms of speech. During the period from fifth to fourth century BCE, the Buddha Gautama adopted Pali for preaching the Buddhist doctrine, and was perhaps not the first of the heterodox teachers to do so; the implication is clear that Prakrit was the preferred form for reaching large audiences, especially those not committed to Brahmanical traditions. By this time, Epic Sanskrit was also developing as a distinct style, with many prakritic features (see above).

The Sanskrit described by Panini in the fifth century BCE was probably still a spoken language: Hock and Pandharipande point to the distinction made by Panini between **bhāṣā** 'speech' (from **bhāṣ** 'speak') and terms like **chandas** or **mantra**, used to refer to the archaic ritual language, as well as certain of Panini's rules which seem to reflect a living language (1976: 115–16). But it seems clear that even this language was restricted to the educated elite, and probably also regionally limited to the area of Aryavarta (OIA **āryāvarta**, see Deshpande 1979b: 4, 49). Only the educated had full access to it, since training was required if one wished to use it correctly: Buddhist sources indicate that young men of the three twice-born castes went to Brahman teachers in the north for instruction in Sanskrit (Hock and Pandharipande 1976: 119).

Thus the question arises whether at this stage Sanskrit was learned natively by anyone. The situation might have been, as Hock and Pandharipande suggest, similar to that of modern German (at least up to the mid-twentieth century), in which most speakers learned a local dialect at home and were taught the standard form in schools, whereas the variety used in the homes of the elite was close to the standard form (1976: 117). It might equally well have been more like the diglossic situation in modern Arabic, in which the classical or older form of the language

is learned exclusively in schools, while local colloquial varieties differ considerably from the classical form as well as from each other (see Ferguson 1964).

The Prakrit inscriptions of King Ashoka in the third century BCE provide a concrete demonstration of the wide currency of Prakrits at that time: clearly it was necessary to use Prakrit in order to reach a wide popular audience. The "dialects" of the Ashokan inscriptions no doubt bear some resemblance to the actual regional dialects of the period, though they were inevitably somewhat stylized and stereotyped (see **6.1**). It has been claimed that, although Ashoka's court used Prakrit for public proclamations, Sanskrit was probably used for internal administrative and diplomatic communication (Hock and Pandharipande 1976: 123); while this may have been true to a limited extent, there is no direct evidence for it.

According to Deshpande, by the time of Katyayana (OIA **kātyāyana**, *c.*300 BCE) "Sanskrit was definitely reduced to the status of a literary, academic and ritual language and was being preserved mainly in the orthodox Vedic ritual and scholastic circles" (1978: 102). Patanjali (**patanjali**, *c.*150 BCE) distinguishes between the **šiṣṭabhāṣā** or language of the learned, which uses correct Sanskrit forms like **kṛṣi** 'farming', and the **lokabhāṣā** or people's language, with prakritic forms like **kasi**. The Kamashastra (OIA **kāmašāstra**) prescribes diglossia in Sanskrit and in the **dešabhāṣā** (local speech, or vernacular) for the cultured man-about-town (Hock and Pandharipande 1976: 114). In contrast to Panini's usage of the term **bhāṣā** 'speech' to refer to Sanskrit, the Natyashastra (OIA **nāṭyašāstra**) (second century BCE, though recording an earlier tradition) uses it in reference to the Prakrits (Deshpande 1979b: 22).

Though Sanskrit later experienced a "renaissance" in which it replaced Prakrits in inscriptions, in secular literature, and even in some of the Buddhist and Jain writings, this "climax of classical Sanskrit literature is reached at a time when the Sanskrit language itself ceased to be anyone's first language" (Hock and Pandharipande 1976: 121). In the post-Mauryan period, Sanskrit was spoken only as a "scholastic second language . . . learned . . . from grammar books", a situation which is "clearly reflected in the discussions in Sanskrit grammar" (Deshpande 1979b: 11).

2.84. Other languages

Early references to specific non-Aryan languages are difficult to find in OIA, though many names of peoples occur. Witzel (1999b) has shown that many of the non-Indo-European names of tribal groups, places, and individuals appear to be derived from Munda or Austro-Asiatic languages (see **3.21–2**). Among references to known languages, terms for 'Dravidian' first occur in post-Vedic texts (e.g. Manu **draviḍa**), the Epics (Mahabharata **dramiḍa**), and Pali (**damiḷa**), but in the early stages at least, these terms seem to refer to the people rather than their language. The same is true of **yavana** 'Ionian, Greek' (later applied to Mohammedans and other foreign groups), which first occurs in Panini. The term **dāsa** apparently referred to a non-Aryan people, but the descriptive term **mṛdhra-vāc** which is applied to them has been translated variously as 'of unintelligible speech' and as 'of

abusive or argumentative speech' (Deshpande 1979b: 2). Even if the first alternative is accepted, it is not fully clear whether reference is to a foreign tongue or to a version of Indo-Aryan which was perceived as mutilated.

The term **mleccha** 'foreigner, barbarian' is related to the verb **mlecchati** 'speaks indistinctly' (or 'speaks an unintelligible language'), which occurs in late Vedic. Panini cites a form **mliṣṭa** 'spoken indistinctly or barbarously' (Monier-Williams 1899 s.v.). According to Thapar, "the word **mleccha** occurs in Later Vedic literature (e.g. Śatapathabrāhmaṇa III.2.1.23–24), and is essentially a term of contempt for those who cannot speak the Aryan language and only gradually comes to acquire the meaning of a barbarian in a cultural sense" (1978: 230).[24] The Asuras (OIA **asura**, OIr **ahura**) are condemned in an oft-cited passage of Patanjali's Mahabhashya for their use of **-l-** for "standard" OIA **-r-**; thus they were regarded as poor speakers of Indo-Aryan, but not necessarily non-Aryans.[25]

These remarks notwithstanding, few scholars of OIA would hesitate to consider those referred to by the terms **dāsa** or **mleccha** as outsiders in a cultural sense – but see Parpola 1988 for the proposal that the Dasas were pre-Vedic Indo-Aryan speaking immigrants from Central Asia. With regard to the Asuras there might be more hesitation because of the connection with Old Iranian (see below). The Dasas, who are mentioned in the early books of the Rigveda, were clearly an alien people. They are referred to as 'black-skinned' (**kṛṣṇa-tvac**) and noseless,[26] but the most prominent fact about them was their alien culture (Thapar 1978: 129). In fact, Dasa may have been a cover term for all those who did not worship the Aryan gods or perform Aryan rituals; they were "godless (**adeva**), nonsacrificers (**ayajyavaḥ**), non-believers in Indra (**anindra**), worshippers of dummy gods (**mūradeva**) and phallic gods (**šišna-deva**)" (Deshpande 1979b: 2). The distinction in the early Rigvedic texts between the **āryavarṇa** or 'Aryan class' and the **dāsavarṇa** or 'Dasa class', in Thapar's opinion, suggests "a rather simple division into 'us' and 'them' where political success justifies the superiority of the former over the latter" (1978: 154). According to Deshpande, the Dasavarna referred to all the indigenous peoples (1979b: 1).[27]

It is probable that at the earliest stage the OIA term Dasa referred to a particular people, about whom some details were known. The names of some Dasa chiefs are recorded (Thapar 1984: 27), as well as the existence of Dasa clans (op. cit. 47). The Dasas are said to live in fortified habitations and to be treacherous (Thapar 1978: 154). Other peoples mentioned in similar contexts are the Dasyus (OIA **dasyu**), noted for "their variant religious beliefs" (Thapar 1984: 44) and the Panis (OIA paṇi), who were given to cattle lifting (Thapar 1984: 24).[28] On the other hand, the situation was clearly not as black-and-white as some have depicted it. Hock notes, for example, that Balbūtha Tarukṣa, presumed to be a Dasa, is mentioned as patron of a Vedic seer. In the descriptions of the famous "Battle of the Ten Kings", there are combatants with Aryan-sounding names (such as Vasiṣṭha and Bharata) and those with non-Aryan-sounding names (such as the Sṛñjayas and Śimyu) on both sides (Hock 1993[1996]: 86–7).

The term **mleccha** occurs often in collocation with **bhāṣā** 'speech' and **deša** 'country, region'. It is probable that in the OIA Brahmanical sources, the lands

designated as **mleccha-deša** included not only areas in which non-Aryan languages were spoken, but also those Indo-Aryan-speaking areas which were regarded as religiously unorthodox. According to Thapar, in later Vedic literature when the geographical focus shifted from the Sapta-sindhu (the Indus Valley and the Panjab) to the Ganga–Yamuna **doāb**, "the northern Panjab and the trans-Indus region came to be regarded as **mleccha-deša**" (Thapar 1984: 159). The land to the east of the Ganga–Yamuna region, including Magadha, which was to be the home of Buddhism and of the Mauryan dynasty, was also considered a **mleccha-deša** at this period (op. cit. 162). Apparently the Brahmans of Aryavarta, which had the Ganga–Yamuna doab as its center,[29] regarded as **mleccha-deša** any region in which Sanskrit and brahmanical rituals were not sanctioned – whether the language of that area was a form of Indo-Aryan or not – whereas to the Buddhists the term meant primarily those lands in which non-Aryan languages were spoken. A number of forms cited in this context are clearly from Dravidian languages.[30]

Another group requiring mention in this connection are the Yadavas (OIA **yādava**). The Yadavas, though regarded as descendants of Bharata, were according to Thapar "clearly not a primarily Indo-Aryan speaking group" (1978: 260). They are associated with cross-cousin marriage in the Puranas, and apparently were found most concentrated in Saurashtra (1978: 349, see also note 69 on p. 360). Because of the wide geographical distribution of those claiming Yadava descent, Thapar believes that some of the lineages associated with the Yadavas were not actual segments of the original Yadava lineage, but "separate groups who were either conquered or else were later integrated into the Yadava lineage in a period when such integrations became a means of acquiring social status" (1978: 342, v. also 343). Specific Yadavas are mentioned in the Rigvedic **dāna-stuti** hymns (composed to praise those who gave generous gifts, 1978: 109).[31]

Thus it is clear that the members of the Old Indo-Aryan speech community were aware, even from early Rigvedic times, of the existence of foreign peoples who spoke other languages. On the other hand, there is no evidence that the OIA speakers were aware of any distinctions among these foreign groups until post-Vedic times. According to Deshpande, "Hatred is the dominant Aryan reaction to non-Aryan languages in the Rgveda" (1979b: 3), though Kuiper (1991) has challenged this view. It seems probable, on the basis of this evidence, that there was a good deal of bilingualism and diglossia in ancient India, with those non-Aryan groups who dealt with the Aryan Brahmans being obliged to learn some form of Indo-Aryan (Sanskrit or Prakrit) for day-to-day communication. On the other hand, the presence of many words of foreign origin in Vedic from the earliest times indicates that this was not a one-sided process.

2.9. Summary

This chapter presents background information on the linguistic situation in South Asia, starting from the earliest period for which useable data are available, and including the current locations of major languages, as well as those minor

languages which are of historical importance. In **2.2** the prehistory and history of Indo-Aryan is traced briefly from the Rigveda through Middle Indo-Aryan (MIA), including the inscriptions of Ashoka in the third century BCE, to the modern (NIA) languages. The subgroupings and chronology of Dravidian and Munda languages are discussed respectively in **2.4** and **2.5**. Other languages, including languages of the Tibeto-Burman family, as well as isolated languages and languages whose existence is inferred from traces found in other languages, are mentioned briefly in **2.6** and **2.7**. The sociolinguistic situation in ancient India is discussed in **2.8**, with a focus on linguistic variation and diglossia within the Indo-Aryan speech community.

Notes

1 General works on South Asian languages and linguistics: Sebeok 1969, Shapiro and Schiffman 1981 (with online bibliography, see Schiffman and Shapiro 2003). For Indo-Aryan languages see Bloch 1965, Masica 1991, Burrow 1973a, Turner 1966, 1975, Mayrhofer 1953b, 1986. For Dravidian see Krishnamurti 2003 (with many references), Burrow & Emeneau 1984, Zvelebil 1970. For Munda see Pinnow 1959, Zide and Zide 1973, 1976.

2 Comparable linguistic areas are found in other parts of the world, such as the northwest coast of the United States and the Balkan region. See Chapter 4 for discussion and references.

3 See Masica 1991: app. I (pp. 421–45) for information on individual NIA languages.

4 See **2.23** regarding the regional groupings shown in Figure 2.1.

5 While the evidence of linguistic structure shows Vedic Sanskrit to be the earliest known form of Indo-Aryan, the hymns were conserved by an exceptionally accurate tradition of oral transmission, and the earliest manuscripts belong to the beginning of the second millennium CE. Note Witzel's comment: "We can actually regard present-day Rgveda-recitation as a *tape recording* of what was first composed and recited some 3,000 years ago (Witzel 1995a: 91; italics in original).

6 Iron, the 'black metal' (OIA **kṛṣṇa ayas, śyāma ayas**), which is mentioned in the Atharvaveda (AV), the second oldest OIA text, was introduced to the region about 1200 BCE (Rau 1974, 1983; Witzel 1995c: 4).

7 References to the pastoral life are much more frequent in the Rigveda than mentions of agriculture. On the other hand, since pastoral nomads depend for their subsistence on the products of agriculture, it is an unresolved question whether the Rigvedic culture depended solely on local farmers, or whether some Indo-Aryan speakers were also (part-time?) agriculturists.

8 See Beekes 1995, Pokorny 1959, and Pedersen 1931 for further information about proto-Indo-European. For scholarly opinions about when and where this speech community existed, see recent issues of the *Journal of Indo-European Studies*.

9 See Witzel 2001 for a thorough debunking of the notion that the Indo-Aryan languages were indigenous to South Asia.

10 "...the early grammarians described final occlusives as repressed or 'weakened', that is, implosives on the same level as occlusives in contact with following occlusives... This evolution reaches finality at the time of the earliest Middle Indian in which even the implosion of the old occlusives...was entirely lost. All finals in Middle Indian are vowels and the new vowels have persisted up to the modern period" (Bloch 1965: 79).

11 See Hock and Pandharipande (1976) and Deshpande (1979b).

12 The original DED of 1961 was supplemented by Dravidian borrowings from Indo-Aryan (DBIA, Emeneau and Burrow 1962), the Dravidian etymological dictionary supplement (DEDS, Burrow and Emeneau 1968), the "Dravidian etymological notes" (DEN, Burrow and Emeneau 1972). These, together with new materials, were incorporated into the revised 1984 edition of the DED, now referred to as DEDR).

13 It has been pointed out by a number of scholars that the present locations of the North Dravidian languages may be the result of migrations, on the basis of their own traditions as well as other evidence (for a recent summary see Hock 1993, with references). However that may be, this tells us little or nothing about their earliest locations.

14 In an earlier paper (Southworth 1995) I accepted a different division of Dravidian, based on work by David McAlpin (1981). I now consider that division to be invalid, as Krishnamurti (2003) has provided new evidence for the validity of the North Dravidian subgroup (including Brahui), and since there seems to be no clear evidence for the existence of a subgroup consisting of South and Central Dravidian (referred to in my earlier paper, and by McAlpin, as PD2) – but see the discussion in **8.22**.

15 Peter Gardner's application of lexicostatistics to Dravidian (Gardner 1980) provides a mean date for the separation of Brahui from the rest of Dravidian of 6100–5000 BP, or 4100–3000 BCE. Gardner's figures show Brahui as more or less equidistant from all the other branches (with a range of 5020–6075 years of separation), a notion which is not supported by the comparative evidence, and which may be a result of the isolated position of Brahui, which has been surrounded by Iranian and Indo-Aryan languages for many centuries and thus has perhaps undergone a higher rate of vocabulary replacement than other Dravidian languages. While one cannot reject Gardner's figures (see the following note), the weaknesses in glottochronological dating have already been pointed out in **1.5**.

16 In Chapter 3 Appendix, I have listed a number of words belonging to the period of Proto-Indo-Iranian, which appear to be of Dravidian origin. If this claim should turn out to be true, then the period of Proto-Dravidian would be pushed back to a considerably earlier period, perhaps as early as 3000 BCE.

17 See Figure 3.1 for the locations of Munda languages.

18 See Shapiro and Schiffman 1981: 107–15 for the classification of these languages.

19 H. D. Sankalia included the Indus Valley or Harappa Civilization, as well as "the various Chalcolithic cultures which were contemporary with and very often immediate successors of the Indus Civilization" (Sankalia 1974: 7–8).

20 This is not the actual sequence of attestation, since the earliest attested writing in Indo-Aryan is the body of Ashokan inscriptions in the third century BCE, and all the Vedic and Sanskrit texts are known only from later manuscripts.

21 See Lee 1986: 158. This word is probably of Dravidian origin; see **3.22A(4)**.

22 Lee takes issue with this argument, claiming that the evidence of the dramas demonstrates coexistence, but not diglossia, since there is no reason to believe that the dramatic usage portrays anything close to reality. She concedes, however, that "The hyper-Sanskritisms . . . are stronger evidence for diglossia" (1986: 160).

23 It is clear that full understanding of the formal variety of Tamil is not essential to the enjoyment of a film or drama, and in fact Tamil informants report enjoying political speeches in high Tamil just because they "sound so nice" (from the author's field notes, 1967).

24 This word is probably of Dravidian origin: see **3.22A(3)**.

25 Thapar 1978: 182, note 18: The Asuras "are described as demons, but also as a maritime people whom the Aryans of the Rigveda had to contend with. Were they the people of the Harappa culture or were they a branch of the Aryans who came from the southern coast of Iran? Archaeological remains in Chota Nagpur are associated with the Asuras." . . . Note also that **-l-** for **-r-** is supposed to be an eastern trait (see **6.16**).

26 Thapar (1978: 154) uses the term 'snub-nosed'; the OIA term, **anās**, can be construed either as **a-nās** 'noseless' or as **an-ās** 'mouthless', that is, without speech (M. Witzel p.c.).

27 Some (including Deshpande 1979: 1) translate varṇa as 'color', presumably relating it to the theory of varnas of the later Vedic literature. It is not clear, however, that this term implied anything about differences in skin-color at this stage.

28 **dāsa** later = 'slave' (Thapar 1984: 42ff., 107); according to Thapar, at a later period they were not regarded as a separate community (1984: 83).

29 Thapar 1978: 182–3, note 21: "ārya-varta was traditionally the region inhabited by the āryas. Its precise geographical area is difficult to define as the concept was not static in history. Broadly speaking, however, the Ganges–Yamuna Doab and the plain of Kurukshetra to the north of Delhi would roughly correspond to ārya-varta, in the strict sense. Some texts extend the definition to include almost the entire Indo-Gangetic plain, e.g., Manu II, 17–74."

30 According to Thapar (1978: 153 and 181, note 8), Buddhaghosha explains **mleccha** as referring to 'andha damil ādi' (Andhra, Tamil, etc.). The Jaimini Dharmashastra lists some (sanskritized) Dravidian words as characteristic of **mleccha** speech, and Panini makes reference to the onomastic suffix **-an** (a Dravidian form) in the names of members of the Andhaka, Vṛṣṇi, and Kuru tribes.

31 One might also mention a group known as the Vratyas (OIA **vrātya**). Thapar states that the Vratyas "remain something of a puzzle. But whether Aryan or non-Aryan, they were looked down upon and held as inferior . . . the *vrātyastoma* is at first an initiation ritual and finally takes on the form of an expiation ritual . . ." (1978: 128). They were not Brahmanical in culture, and though they spoke the language of the initiated with difficulty, they were not dismissed as **mleccha**; thus they were clearly a powerful group, and some "tribes of indigenous origin" (including the Drāviḍa, Śabara, Kirāta, Mālava, Śibi, Trigarta, and Yaudheya) are referred to in some sources as "Vratya kshatriyas", and there is "evidence from numismatic sources of the increasing political importance of some of these groups" (1978: 165). Falk (1986) considers the Vratyas to have been originally renegades from Brahman families.

3

PREHISTORIC LANGUAGES OF SOUTH ASIA

Contents

3.1. Summary of previous work

Non-Indo-European elements in OIA have been of interest to scholars since the early years of Indo-European studies. The subcontinent was clearly not a linguistic vacuum at the time the first OIA speakers entered it. On the other hand, the earliest linguistic records relating to the subcontinent are in Indo-Aryan languages (see **2.21**), and give very little direct information on the presence of other languages (see **2.84**). Thus any linguistic elements of non-Indo-European origin in OIA can provide clues about the kinds of contact which took place between the OIA speakers and other peoples in the region, just as (e.g.) borrowings between Proto-Indo-Iranian and Finno-Ugric provide evidence of prehistoric contact between speakers of these two groups during the period of common Indo-European (Burrow 1973a: 23–7). Information about prehistoric languages in South Asia is derived almost entirely from the presumed traces – both loanwords and structural changes – which these languages have left in Old and Middle Indo-Aryan.

Over the years scholars from many countries have hoped to shed light on the early linguistic situation in South Asia through the decipherment of the Indus Valley seal inscriptions. Sadly, in spite of the expenditure of much erudition and ingenuity on these inscriptions – and several claims of actual decipherment – there has been as yet no generally accepted interpretation of the script. A number of serious attempts to interpret the script as Dravidian have been made, notably by Fairservis (1992) and Parpola (1994), both of whom made important contributions to knowledge of the technological and cultural background of the inscriptions. Parpola's work goes far beyond the script itself, with an excellent section on the Linguistic Context (pp. 123–76), including a discussion of Brahui. The best-known attempt to interpret the script as Indo-Aryan is by S. R. Rao (1982) – see the critique by Mahadevan (1981–82). Possehl (1996) has reviewed all of the attempts at decipherment up to 1996 and discussed both positive and negative aspects of each.

This problem has now received an unexpected solution. Farmer *et al*. (2004) throw into doubt the whole thesis that Indus symbols encoded speech. Their argument is based in part on the brevity of the inscriptions, which (averaging under five signs each) are far shorter than those produced by any known script that left behind so many inscriptions (4,000–5,000 are known). Since the 1920s, the brevity problem has been bypassed by claiming the Harappans wrote on perishable materials; but this is contradicted by the absence of all known markers of manuscript production. Finally, the nonlinguistic nature of the signs is confirmed by extremely low sign-repetition rates in single inscriptions, as compared to known scripts, even though a few high-frequency signs dominate in the inscriptions. Whatever the sense of Indus symbols, they are not likely to provide clues about the languages of the time; this leaves us with the linguistic clues in Old Indo-Aryan, which are the subject of this chapter.[1]

Investigations of foreign elements in OIA have focused mainly on the Austro-Asiatic and Dravidian language families as potential sources, with the greater amount of attention given to Dravidian until very recently. Dravidian lexical borrowings, or loanwords, in Sanskrit have been discussed by Burrow (e.g. 1945, 1947a,b, 1973a: 380–7), Emeneau (e.g. 1943, 1954), Kuiper (1955), and Southworth (1979b), among others. OIA loanwords in Dravidian languages have been treated in Emeneau and Burrow (1962). According to Burrow, the borrowing of Dravidian words into Old Indo-Aryan occurred mainly "between the late Vedic period and the formation of the classical language" (1973a: 386). Witzel (1999b), however, believes that Dravidian borrowings occur from the middle Rigvedic period on (see **3.22**). Kuiper (1967) and Emeneau (see Dil 1980) have made important contributions to the study of structural similarities between OIA and Dravidian (see **3.32**). Hock (1975) presents a critique of much of this work. Austro–Asiatic (including Munda) words in OIA have been discussed by Lévi (1923), Przyluski (1926), Kuiper (1948, 1955, 1991), Mayrhofer (1951, 1953a), and most recently Witzel (1999b). Several of these authors, as well as Hock (1975: 3.1), have noted that the earliest foreign elements found in the RV are Austro–Asiatic in appearance. (See further discussion in **3.21**.)

The dental–retroflex distinction in OIA consonants (dental **t d n l s** vs retroflex **ṭ ḍ ṇ ḷ ṣ**) has been attributed by many to Dravidian. The earlier history of the controversy is summarized in Kuiper (1967), which takes the view that the feature is of Dravidian origin. Several scholars (see e.g. Hock 1975) oppose this view, claiming that there is no evidence that the development of a retroflex–dental contrast in OIA was anything other than a purely internal innovation. Deshpande (1978, 1979a) attempts to show that this was a post-Rigvedic phenomenon, which was not present in the earliest (unattested) form of the Rigveda. Witzel argues that words with retroflex consonants in the early Rigveda are probably from a pre-Mundaic language of Panjab (1999b: 12); see **3.31**.

The evidence of loanwords in Old Indo-Aryan is analyzed in Section **3.2**. Evidence for syntactic influence of Dravidian languages on OIA has been presented in a well-known paper by Kuiper (1967) and in several papers by Emeneau (see Dil 1980). Hock 1975 has attempted to refute this evidence as far as the Rigvedic period is concerned. Witzel (1999b) and Tikkanen (1988) have discussed other possible sources of structural innovations in OIA. This evidence is discussed in **3.32**.

3.2. Prehistoric languages of South Asia

Since the Old Indo-Aryan texts date back to the second millennium BCE, and since languages of other families are not directly attested until the CE (possibly a few centuries before in the case of Old Tamil), the evidence for the presence of other languages depends on the analysis of non-Indo-Aryan linguistic elements in OIA. Figure 3.1 presents a linguistic map of pre-Indo-Aryan South Asia as reconstructed from this evidence. (Also included on the map are the approximate locations of the existing Dravidian and Munda languages.) It must be emphasized that the locations depicted here are very approximate, and even the existence of some of these languages is controversial. Note that this map covers a period of several millennia, including the probable earliest locations of Munda/AA and Dravidian languages in the subcontinent, the inferred locations of languages such as "Meluhhan" and "Proto-Bhili", and the modern locations of some languages such as Brahui, Malto, and Nahali, which are not necessarily their ancient locations. The languages which are smaller in extent, such as Burushaski, "Proto-Bhili", and Nahali, may possibly be remnants of larger groups of languages, as seems to be the case with the Munda languages.

Three types of languages are shown in Figure 3.1:

(1) languages belonging to known families: (a) Munda and related Austro-Asiatic languages, (b) Dravidian languages, and (c) Tibeto-Burman languages;
(2) isolated languages of no known language families;
(3) languages whose prior existence is inferred from traces left in existing languages.

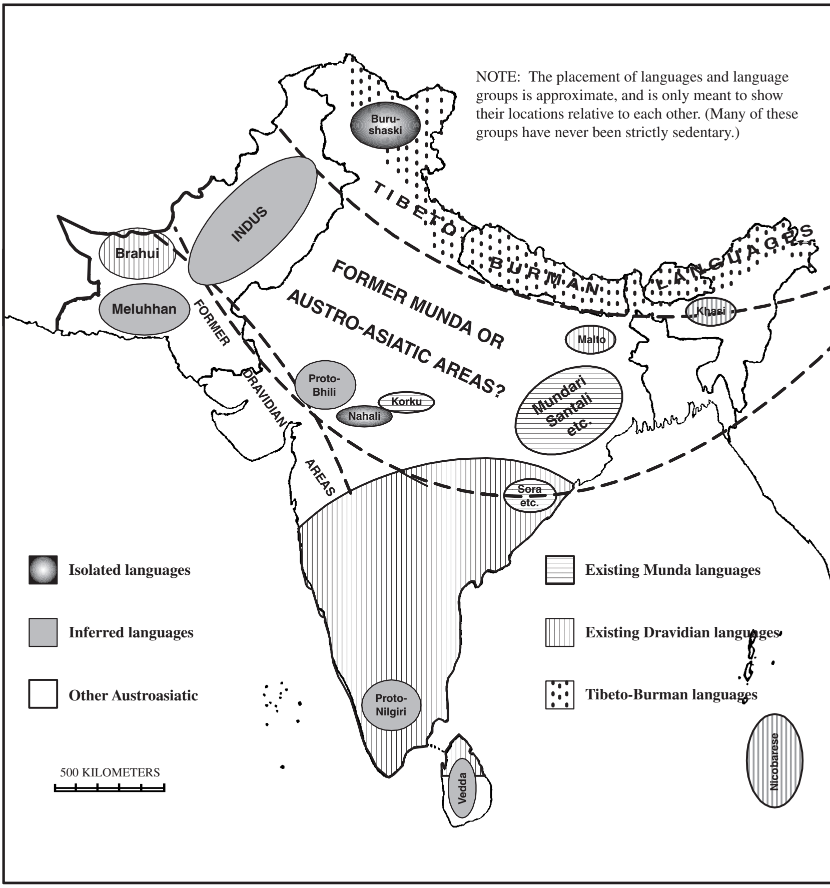


Figure 3.1 Pre-Indo-Aryan substratum languages.

These types are given in detail in the following list:

(1a) *Munda/Austro-Asiatic languages*. The prehistoric distribution shown in Figure 3.1 is based on (a) the locations of existing Munda languages, (b) the existence of probable Munda/Austro-Asiatic words in OIA throughout the entire Vedic period, and (c) the generally accepted genetic link between Munda and other Austro-Asiatic languages in South and Southeast Asia.

(1b) *Dravidian languages*. See **2.4** for a brief discussion of Dravidian prehistory; see **8.4** for further details.

(1c) *Tibeto-Burman languages*. Various languages of this family are found in the foothills of the Himalayas. In general, very little is known about their earlier history. Witzel mentions Tibeto-Burman languages in the Himalayas and in the area of Avadh (around the city of Lucknow) in Uttar Pradesh: the name

of the latter area in Vedic, Kosala, along with the river Kosi in Bihar, is probably from a Tibeto-Burman source. (See Witzel 1999b: 63ff. for this and several other probable cases.) Newari or another Tibeto-Burman language may have been spoken in the area of northern Bihar where the Buddha spent most of his life, and may have played a role in the (alleged) merger of **l** and **r** which was characteristic of the Prakrit inscriptions of the eastern region (see **6.14B**).

(2) *Isolated languages*. Those shown in Figure 3.1 are among a handful of languages with no known genetic links to other languages, inside or outside of South Asia: Burushaski and its predecessors in the northwest, which may be the source of a few words such as **kīḷāla** RV 'biestings, a sweet drink' (Bur. **kilāy**) (Witzel 1999b: 52); predecessors of isolated languages scattered about the subcontinent, including Kusunda in Central Nepal, Tharu in Southern Nepal and UP, and Nahali in Central India. These languages may belong to the oldest surviving linguistic strata in South Asia, though there is no way of knowing if they are truly indigenous to the subcontinent. One or more of them may have previously occupied larger areas, but at present we know nothing of their earlier history.

(3) *Inferred languages*. These are languages whose existence is inferred from traces (vocabulary and/or grammatical constructions) found in existing languages. Their prehistoric status is comparable to that of the isolated languages.

(a) The "Indus" language(s), which served as the source of numerous words, mainly names of plants, found in OIA and early Dravidian (see **3.23**, also Southworth 1988);
(b) "Meluhhan," the source of some 40 "Indian" words found in ancient Mesopotamian sources, referring to trade goods originating in the Indus Valley. This language may have been located in the hilly areas of Baluchistan, near to the Indus Valley (Witzel 1999b: 37–8);
(c) An unknown substrate language, or group of languages, in the area of Bhili, Ahirani, Ḍangi, and Katkari (the region where Gujarat, Rajasthan, Madhya Pradesh, and Maharashtra adjoin each other) which has left its mark on the lexicon, and perhaps the grammatical structure, of these languages. I have provisionally dubbed this substrate "proto-Bhili";[2]
(d) "Proto-Nilgiri," a pre-Dravidian substrate in the Nilgiris in South India (Zvelebil 1990: 63–70; Witzel 1999b: 4);
(e) The Vedda substrate in Srilanka, inferred on the basis of loanwords and collocations in Sinhala (De Silva 1972);
(f) Witzel also refers to "Central Asian substrate(s)" – not shown here – which may be the source of a number of words in early OIA as well as various Iranian languages (1999b: 8): for example, OIA **usṭra** 'camel', **khara** 'donkey' (see **3.23A1**), **isṭi** 'brick' (Witzel 1999b: 79–81);
(g) Masica (1979) posited a "Language X" to account for agricultural words of unknown origin in Hindi–Urdu. Though Masica started with Hindi vocabulary in tracing the history of these words, the large majority of them

> are of general occurrence in Indo-Aryan. On the assumption that the ancestors of all Indo-Aryan languages passed through the Indus Valley during the OIA period, a source in that area seems most probable, and in the absence of evidence pointing to some other specific location, it seems reasonable to posit the "Indus" languages as the source of this material.[3] Of course, this language (group) may not have been confined to the Indus Valley region.

Witzel's article "Substrate languages in Old Indo-Aryan" (Witzel 1999b) uses the evidence of the Vedic texts to identify the various substrate languages by time and place, as well as by phonological and morphological features of their language. Witzel notes that the RV contains some 383 words ("roughly 4 percent of its hieratic vocabulary")[4] which are not of Indo-Aryan or Indo-European origin, as shown by their phonological and morphological structure, including a number of place names in Panjab and the Ganga–Yamuna doab. Witzel believes that the oldest stratum of these loanwords in the Rigveda is derived from Munda or related (and otherwise unknown) Austro-Asiatic languages which he designates as "Para-Munda." See **3.21** for a discussion of this evidence. Dravidian borrowings, according to Witzel, do not appear until the middle Rigvedic period (but see **3.22B**).[5] Section **3.23** discusses loanwords from other sources. The largest number of words in this group are those which occur in OIA and in early Dravidian, and which do not appear to be original in either group; since they do not have the typical Munda characteristics, they may be provisionally considered to have originated in the otherwise unknown "Indus" languages.

3.21. Munda and Austro-Asiatic (AA)

As noted in **2.5**, Munda languages form a branch of the Austro-Asiatic (AA) language family, which includes various languages of Southeast Asia as well as Khasi (a branch of Mon–Khmer spoken in Assam) and Nicobarese (in the Nicobar Islands in the Bay of Bengal). The AA languages may in turn be a branch of the larger Austric family, which includes Malayo-Polynesian (see Blust 1996a). Munda languages (Figure 2.5) are at present located mainly in eastern India, the westernmost language being Korku on the middle Narbada. However, Kuiper has long held that the earliest identifiable foreign words in the Rigveda are of Munda or AA origin (1948, 1955, 1991), which would imply the presence of speakers of these languages in the Panjab as early as the second millennium BCE. The Rigveda alone contains more than 300 such words.

There are as yet few definitive Munda or AA etymologies for these words, and of those that do exist, some still require rather tortuous argumentation. Scholars involved in this endeavor have been hampered by the lack of an etymological dictionary, though that task has been undertaken now by David Stampe and others (Stampe 2003). Many of the proposed Munda/AA words are names (of individuals, tribal groups, or geographical features). It is of course not certain that these names were transmitted directly into OIA by Munda speakers, though as Hans

Hock has pointed out, the geographical locations provided for these names "form a chain extending from the eastern border of Kashmir to the Central Mountains where even today, Munda-speaking tribes can be found" (Hock 1975: 3.1). Some of the earlier work on Munda borrowings in OIA, for example that of Przyluski, was difficult to evaluate because of the many assumptions, both linguistic and non-linguistic, required to make the derivations plausible.[6] In addition, some of Kuiper's Munda interpretations have been challenged, on the grounds that they may be Dravidian compound words rather than prefixed Munda/AA words (Krishnamurti 2003: §1.7.1).

Nevertheless, the hypothesis of Munda/AA origin for numerous OIA words, both Rigvedic and later, is supported by two important facts: (1) many of these words contain prefixes of types found in Munda and other AA languages, especially double prefixes (see examples given below), and (2) none of the other known language families of the area, including Dravidian, Tibeto-Burman, and Burushaski, use such prefixes. Therefore, in a sense by default, Munda/AA languages become the most likely candidates as sources for these words. Kuiper and Witzel use the term 'Para-Munda' to refer to an unknown Munda or western AA language (or languages) which may have been the source of these words.

There is no need to discuss these etymologies in great detail here, as they can be found in Witzel (1999b), as well as Kuiper (1991); Witzel plans to deal with them in greater detail in a forthcoming monograph. A few cases are mentioned here to show the nature of the evidence. Words with AA-like prefixes which have no Indo-European etymologies are considered to be Para-Munda; for example:

- with the prefix **ka/kā-**: **kakardu** 'wooden stick', **kākambīra** 'a tree';
- with the prefix **ki/kī-**: **kimīdin** 'a demon', **kīkaṭa** 'name of a tribe' (here **kī** may be an old AA plural prefix; the root **kaṭ-** is possibly connected with Santali **kaṭ** 'fierce' or **kaṭkom** 'crab' (as totem);
- with the prefix **ku-**: **kumāra** 'young man' (cf. Munda **məndra**, **mər** 'man');
- with double prefixes (**kə-r-**, **jə-r-**, **kə-n-**, etc.): **ja-r-tila** 'wild sesame' **śa-l-mali** 'a tree, *Salmalia malabarica*', **ki-l-bisa** 'evil action'.

A number of proper names come from "areas where Indus people are to be expected: … eastern Panjab, in Haryana (Kuruksetra), and … well into the Gangetic plains" (Witzel 1999b: 16). These include the aforementioned **Kīkaṭa** tribe and their chieftain **Pramaganda**, both words which show "clear indications of Mundic character," as well as other place/people names such as **Gāndhāra** and **Kamboja**, and river names containing the Munda element **ga(n)d/ga(ṇ)ḍ** 'water, river' (1999b: 16ff.). One important word which seems to have a clear AA source is **lāngala** 'plough' (RV 4.57.4, a late hymn); cf. Santali **nahel**, Khasi **lynkor** [lənkor] < ***lenkol**, Khmer **ankal**; cf. also the Austronesian forms, Malay **tengala**, Makassar **nankala** (Bagchi 1975: 9). See **3.23A(1)**, Witzel 1999b: 25. Note also OIA **kadala** 'banana, plantain' (Epic) (see **7.14(D2)** for background).

On the basis of the totality of evidence, Witzel claims: "In short, a strong Austro-Asiatic substrate is found both in the early Panjab (RV, *c.*1500 BC) as well as later on in the Ganges valley (YV samhitās, brāhmaṇas, *c.*1200–500 BC), a fact that can also be shown in the names prevailing in these areas" (2001: 56). Thus we are probably dealing, during the whole Vedic period, with vocabulary derived from a *non-Dravidian* source, which precedes the appearance in OIA of items of Dravidian origin. As Witzel puts it, "…the Panjab is an area of a Pre-Rgvedic, largely Para-Munda substrate that apparently overlays a still older local level [referred to in this work as the 'Indus' language(s)] which may be identical with Masica's 'language X'…" (1999b: 19).

3.22. Dravidian

The subject of Dravidian loanwords in OIA has been fraught with controversy from the early days of comparative linguistics, particularly for the Rigvedic period. A number of scholars have been reluctant to accept the possibility of Dravidian loanwords in the Rigveda, especially the earliest portion. Mayrhofer, for example, in his etymological dictionaries of OIA (1953b, 1986), has at several points cited Rigvedic attestation as a reason for rejecting Dravidian derivation of a word (see e.g. **phala** in **3.22A(1)**, **khara** in **3.23(1)**), while in other cases accepting Dravidian origin *in spite of* the early attestation: see for example, **kīri**, **kula** in **3.22A(1)**.[7] Reacting against Dravidian etymologies proposed by Burrow (1955), Thieme (1955) proposed a number of Indo-European etymologies for OIA words, some involving complicated and unnecessary assumptions: see **khala (3.22A(1)**, **śava 3.23A(3)**, **anala 3.22A(4)**, **ulūkhala 3.23(2)**. The reasoning behind this reluctance to accept Dravidian etymologies has rarely if ever been made explicit. It is not the case that the language of this period is in any sense 'pure', that is, free of foreign influence, since (as noted earlier) there are over 300 words in the Rigveda alone which are generally agreed to be of non-Indo–European origin.

Witzel (1999b) has attempted to present an "impartial historical relief to the ongoing discussion" in the hope of achieving some consensus on this question. After examining all of the proposed Dravidian loanwords in the Rigveda, he concludes that plausible cases occur only from the middle Rigvedic period onward, and in much smaller numbers than is the case for those thought to be of Munda/AA origin.[8] For the middle and later RV, he accepts as possible Dravidian loans the following words (those discussed here are indicated with references to following sections):

- in middle RV: **phala** 'fruit' **3.22A**(1), **phāla** 'ploughshare' **3.22A**(1), **piṇḍa** 'ball, dumpling', **kuṇāru** 'lame in the arm', **kāṇa** 'blind in one eye' **3.22A**(1), **kulpha** 'ankle', **daṇḍa** 'stick' **3.22A**(1), **kuṇḍa** 'pot, pit' **3.22A**(1), **naḷa** 'reed', **mayūra** 'peacock' (Appendix, item 5);
- in late RV: **ulūkhala** 'mortar' **3.23A**(2), **vriś** 'finger', **bila** 'hole, cave' **3.22**(1), **kuṭa** 'hammer' **3.23A**, **kaṭu** 'bitter' **3.22A**(1), **bala** 'force'.

Sections **3.22A** and **3.23** include a number of additional items from RV and other Vedic sources; some of them are presented here for the first time, while others involve new arguments for etymologies previously proposed by various scholars. Ultimately our goal is to achieve some sort of consensus – if not on every individual item, at least on the time periods in which loanwords entered the OIA language from each source.

3.22A. Probable Dravidian borrowings in OIA

This section contains words appearing in OIA at various stages whose most probable source appears, in the author's opinion, to be Dravidian, including a number of words published here for the first time. As many of these words have been treated in other sources, the following list makes no claim to be exhaustive; the cited sources may be consulted for additional items. All of these items conform to the following conditions (see Burrow 1946: 13–18 and Emeneau 1954: 287–8 for comparable criteria):

(1) The OIA word has no Indo-European etymology whose probability is equal to that proposed for Dravidian.
(2) The Dravidian source word is of comparable antiquity with the OIA word, except where other factors must be invoked, as in the case of **nagara** in **3.22A (3)**.[9]
(3) The source word is etymologizable in Dravidian, or shows its Dravidian origin by phonological developments, variant forms, or semantic developments (see e.g. **candana** in **3.22A (4)**).

Proto-Dravidian reconstructions cited in italics are from Krishnamurti (2003 §1.2.2); others are by the author. Note that Krishnamurti and the author differ as to the level of Proto-Dravidian to which some of these reconstructed forms belong: see **8.23**.

Phonological assumptions

(1) Dravidian /**ẓ**/ has no exact equivalent in Indo-Aryan. It may appear as /**r**/ or /**l**/, or even /**ḍ**/ or /**ḷ**/. For examples see **kīri**, **kula**, **phala** in **3.22A (1)** below, **karīra**, **kāla** in **3.22A (4)**.
(2) Early stages of Dravidian had no distinction between voiced and voiceless consonants. Stops were presumably phonetically voiced in non-initial positions except when doubled (Zvelebil 1970: 76ff.). Thus one would expect initial Dravidian stops to appear as voiceless stops in OIA, though voiced stops might appear occasionally, especially if the item in question frequently appeared non-initially in phrases. See **daṇḍa** in **3.22A(1)**, **goṇī** in **3.22B(5)**, **gaṇḍa** in **3.22A (3)**.

(3) The aspirated/unaspirated distinction did not exist in early Dravidian, and even now occurs mainly in writing and in formal styles of speech in Telugu, Kannada, and Malayalam. In the large majority of cases, Dravidian stops appear as unaspirated stops in Indo-Aryan, though there are a few cases of aspirated stops, frequently in word-initial position. See **phala**, **khala** in **3.22A (1)** later.
(4) In cases like OIA **mukta** 'pearl' ← PSD ***mutt*** (**3.22A**(**4**)), we must assume that there was an awareness among OIA speakers of the relationships between OIA words with C_1C_2 and Prakrit (or non-IA) words with C_1C_1 (or C_2C_2). (**See 2.2** and the discussion of diglossia in **2.83**.) While some of these cases presumably involved Prakritic assimilation of original OIA forms with C_1C_2, it is generally agreed that there were also numerous cases of the reverse, that is, 'Sanskritization' or folk-etymologizing of the original form with C_1C_1.
(5) See Southworth 1979b: 201–2 for equivalences within Dravidian.

Burrow (1971) has pointed out a number of OIA doublets which differ only in that one has a dental consonant where the other has a retroflex, for example OIA **pattana**/**paṭṭaṇa** 'city, town' (**3.22A**(**4**)). In a number of these cases, including some with clear Indo-European etymologies, it is the form with dental which has been shown to be older. Previously it had been believed that such cases required an explanation in terms of "Fortunatov's Law", that is, **l** + dental → retroflex, or in terms of an earlier **r** or **ṛ** which later disappeared – or failing that, must be regarded as indicating borrowing (or "influence") from indigenous languages. However, Burrow now proposes to explain these cases in terms of a spontaneous "fission" of the inherited dentals into dentals and retroflexes (Burrow 1971). Clearly, spontaneous "fission" is no explanation for a change, and such an assumption runs counter to the assumption of the regularity of sound change (**1.21C**). In examining Burrow's claim, we must note first of all that some of his cases have quite plausible, and in some cases probable, Dravidian etymologies. For example, **pattana**/**paṭṭaṇa** 'city, town' (the dental is attested in MBh, the retroflex later) can be plausibly derived from PSD ***paṭṭi(ṇam)** 'cowshed, settlement, village, town, city' (**3.22A4**). Such cases show the possibility of reversing Burrow's argument entirely. It is easily understandable that a language without the retroflex–dental distinction would substitute dentals for retroflexes in borrowed words in the early stages of contact. Furthermore, if a word like OIA **piṇḍa** 'body' had been borrowed into early Iranian, or proto-Indo-Iranian, we would naturally expect to find it with a dental, since Iranian never experienced the "fission" phenomenon (see M53 s.v.).

Thus the presence of Iranian cognates by no means excludes the possibility of borrowing from Dravidian or other indigenous languageṣ. The possibility of contact between Dravidian and Proto-Indo-Iranian has been denied by most historical linguists who have considered this matter. Such contact would have presumably taken place outside of the South Asian subcontinent proper, in Afghanistan or Central Asia, and if there is evidence for it, it would support the

notion that Dravidian speakers came from elsewhere. The etymological evidence which bears on this point is presented in the Appendix to this chapter, as the arguments involved are rather complex and may not interest all readers.

(1) Rigvedic period

anu 'non-Aryan man' RV; **ānava** 'the Anavas' MBh ← PD ***āṇu/āṉu** 'man' DEDR 399; cf. also the SD1 masc. sing. suffix **-aṉ** as in **avaṉ** 'he' (that one), **ivaṉ** 'he' (this one), and the **-āṉ/aṉ** which occurs as a verbal suffix of the third person singular, e.g. as in Tamil **pōv-āṉ** 'he will go' (Andronov 1969:142ff.). M53: prob. non-Aryan tribal name; M86: not clear.

kaṭu(ka) 'bitter' RV ← PD ***kaṭ-u** 'pain, sting, be pungent, etc.' DEDR 1135, cf. PD ***kaṭ** 'pungent' DEDR 1492. M53, M86: not satisfactorily explained.

kāṇa 'blind in one eye' RV ← PD ***kāṇ-a** 'not seeing' = PD ***kāṇ** 'see' DEDR 1443 + **-a-** neg. suff. (Burrow 1946, Emeneau 1954 fn. 22). M86: unclear, prob. an earlier Prakritism.

kīri 'small, humble, poor' RV ← PD ***kīẓ** 'below, inferior, humble' DEDR 1619. M53: Dravidian origin possible in spite of Vedic attestation; M86: etymology depends on clarification of meaning.

kuṇḍa 'pot, hole, pit' RV ← PD ***kuṇṭ-a-m** 'cavity, pit' DEDR 1669, derivable from PD ***kuẓ** 'pit, cavity' DEDR 1818 (cf. Ta. **kuṇṭa** [= **kuẓ** + **-nt-** 'past' + **-a** 'rel. pple.'] 'hollowed out'); Burrow 1946, 1973a. M53: prob. Dravidian; M86: unclear, perhaps loanword. See **nirguṇḍi** in **3.22A(4)**.

kula 'herd, flock, lineage' RV ← PD ***kuẓ-u** 'assembly, flock, herd, heap' DEDR 1821 (Burrow 1946). M53: Burrow's derivation very plausible in spite of the word's great antiquity; M86 cites possible Iranian cognates.

khala 'threshing floor' RV ← PD ****kaḷa-m/n*** 'open space, threshing floor' DEDR 1376 (Burrow 1946, 1947a, Emeneau 1954). Thieme's interpretation of **khala** as a vernacular form of "the educated form **khara**, used in the sacrificial language as a designation of the square, slightly elevated and specially prepared 'ground where the sacrificial vessels are kept when not in use'" (1955: 439) seems unnecessary. (See C2. OIA **ulūkhala**.) M53, M86: unexplained.

daṇḍa 'stick, club' RV ← PD ***taṇṭa** 'stalk, stem' DEDR 3056; cf. PD ***taṇṭa** '(fore)arm' DEDR 3048 (Burrow 1946). M53 notes the longstanding debate among Indo-Europeanists regarding derivation from an OIA ***dandra** 'tree' ← PIE ***dendro** (cf. Gk. **dendron**), and also states that the arguments for non-Aryan origin deserve attention; M86: unclear. This item is included here because of the greater semantic and phonological plausibility of Dravidian derivation.

phala 'fruit' RV ← PD ***paẓam** 'ripe fruit', ***paẓ** 'ripen' DEDR 4004 (Burrow 1946; Emeneau 1954). Because of the derivation in Dravidian and the considerable variation among the Dravidian languages, M53 allows the possibility of Dravidian origin in spite of Vedic attestation; M86 considers Dravidian origin of this early word to be improbable, even on semantic grounds ("auch

von der Semantik her”). Cf. **pippala** RV ‘berry’, **pippalī** ‘long pepper’, probably from the same source (cf. **ph** in Pali **pipphala/pipphalī**, T8205).

phāla ‘ploughshare’ RV 4.57.8: this word, of unknown origin, has cognates in modern Iranian languages: Farsi **supār**, Pashto **spāra**, suggesting a possible proto-IIr. form ***spār**. Turner (CDIAL9072) and others (M53, M86, Witzel 1999b: 22) suggest possible Dravidian or Munda influence on the form of the OIA word. Cf. PD ***par** ‘split, cut, separate, plough for the first time, etc.’ DEDR 3962.

bila ‘hole, cave’ RV ← PD ***viḷ** ‘open out, open up’ (cf. Ma. viḷḷu ‘crack, aperture’) DEDR 5432 (Burrow 1946). M53: not satisfactorily explained; indigenous origin possible; M86: not clear.

yādava ‘name of a people’ RV ← PD ***yāṭu-v-an̲** ‘goat/sheep herder’ = PD **yāṭu* ‘goat/sheep’ (see **eḍa(ka)** ‘sheep’ in **3.22A**(4)) + ***āṇu/ān̲u** ‘man’ DEDR 399 (see earlier text). M53 lists **Yadu** ‘presumed ancestor of the Yadavas’ as of unknown origin (with a cross-reference to OIA **yātu** ‘magic’); M86: not clear. See above for a discussion of the phonology.

vaṃśa ‘bamboo’ RV, **vañjula** ‘name of various trees and plants including *Jonesia asoka* and *Calamus rotang*’ MBh ← PD ***vañc(-i)** ← PD ***vānk** ‘to bend’; note Te **vampu/vancu** ‘to bend’, **vampu** ‘bend, crookedness’ DEDR 5335; Ta-Ma **vañci** ‘bamboo’ DEDR 5216, PSD1 ***vam-pu** ‘bamboo’ DEDR 5253. M53, M86: no certain connections outside of Indo-Iranian frontier languages: Waig **wãs**, etc.

vṛścika ‘scorpion’ RV (cf. Pali **vicchika**, Pkt **vicchia**, **viṃchia**, **viccu**, **vimcua**, etc.) ← PD ***vī(n)c** ‘throw (as weapon), cast (as net), take aim; quickness, rapidity’ DEDR 5450: note Ma. **viśa** ‘spring-trap, snare for birds, lever’, Te. **bisa** ‘spring, catch’, **bisi** ‘tension’. In the absence of any clear Indo-European etymology, derivation from this Dravidian word which describes the attacking movement of the scorpion, a denizen of the subcontinent, seems most plausible. The OIA word is likely to be a hypersanskritization of an early Prakritic form **vicci/viccu**, possibly through a folk-etymological link to **vṛścati** ‘cleaves, cuts’. M53: uncertain; M86: difficult.

syāla ‘wife’s brother’ RV (also **śyāla**) ← PD ***caḷ/cāḷ** ‘male cross-cousin, wife’s brother, sister’s husband; daughter’s husband’ DEDR 2410; probably Uralo-Dravidian: cf. Finnish **käly** ‘husband’s sister, wife’s sister’, Selkup Samoyed **śāl** ‘brother-in-law’ (Tyler 1968: 811; Collinder 1977: 43). M53, while noting Hoffmann’s argument that OIA could not in principle have an inherited PIE word for the wife’s sister, connects OIA **syāla** to Slavic words in **-r-**. Borrowing of the PD word from an “**-l-**Prakrit” of the Vedic period is more probable. M86: ditto.

(2) Other Saṃhita Vedic

kāṇḍa ‘single joint of a plant’ AV, ‘arrow’ MBh ← PD ***kaṇ** ‘joint in cane or bamboo’ DEDR 1160, PSD ***kaṇ** ‘arrow’ DEDR 1166 (Burrow 1947b). M53: prob. Dravidian; M86: unclear, prob. of MIA or non-IE origin.

taṇḍula ‘grain, esp. rice’ AV ← PD ***ta(ṇ)ṭ-uḷa** ‘beat-en’ (i.e. hulled rice) ← PD ***taṭ** ‘to beat’ (Burrow 1971: 545). M53: unclear; Bloch’s Dravidian derivation shows important semantic parallels. M86: unclear.

paṇḍa(ka) ‘eunuch, effeminate man’ MaitrS, **paṇḍaga** AV ← PD ***peṇḍ** ‘female’, PD ***peṭṭ** ‘female’ DEDR 4395 (Burrow 1945). M53: non-Aryan; M86: not convincingly explained.

bilva ‘wood apple tree’ (*Aegle marmelos*) AV ← PD ***viḷ(avu)** DEDR 5509 (Burrow 1945). M53: Dravidian probably primary because of the number of variants; M86: indigenous tree name?

(3) Later Vedic

udumbara ‘tree or fruit of *Ficus glomerata*’ TS, ŚBr ← PD ***uttu-mara** (← PD ***uttu** ‘date’ DEDR 620 + PD ***mara** ‘tree’ DEDR 4711. In spite of the difference between ‘fig’ and ‘date’, this etymology is to be preferred over Przyluski’s tortuous derivation from Austro-Asiatic words for ‘gourd’ (v. Bagchi 1975: 158–60 and **3.22**), which has been accepted by M53 and others. M86: unexplained.

kuṭumba ‘household’ ChUp ← OIA **kuṭi** (see A51) + SD1 **umb** ‘descendant, child’ DEDR 639. M53/M86: perhaps from Drav. along with **kuṭi**.

gaṇḍa ‘goitre, excrescence, boil’ AitBr ← PD ***kaṭṭ** (cf. Ka **gaḍḍe**) ‘mass, lump’ (Burrow 1947a). M53: prob. Dravidian; M86: unclear.

cūḍā ‘protuberance on brick’ ŚB, **coḍa** ‘topknot’ TS ← PD ***cūṭ(ṭ)i** ‘tuft, topknot’ DEDR 2721 (Burrow 1947b). M53: Dravidian origin most likely; M86: not clear.

nagara, nagarī ‘town’ Ār. Mn. ← PSD ***nakar** ‘house, palace, town, city’ DEDR 3568 (Burrow 1945). In spite of the late attestation of the Dravidian word, the semantic development evident within Dravidian suggests that it is earlier there, though perhaps ultimately of Indus origin. M53, M86: presumably Dravidian. (See **paṭṭaṇa** ‘town’ in **3.22A(4)**, and discussion in **8.31**.) According to Witzel, “ … may be a loan from the southern Indus language or one from the Malwa area” (1999b: 29). The first occurrence of this word in OIA is in the Jaiminiya Brāhmaṇa, a text which Witzel (1989, 1995a) has located between the lower Yamuna River and the Gulf of Cambay, possibly in the vicinity of Ujjain (M. Witzel p.c.).

pañcāla ‘name of a people’ ŚB (cf. **pañcajanā** ‘the 5 peoples of the Rigveda’) ← OIA **pañca** ‘5’ + PD ***āḷ** ‘man, person, adult human’ DEDR 399. Compare OIA **caṇḍāla** ‘name of a people’ for the same suffix. M53/M86: not clear.

mleccha ‘barbarian, foreigner, non-Aryan, stammerer’ ŚB; **mlecchati** ‘speaks an ununderstandable language’ ŚB (cf. Pali **milakkha**, Pkt **miliccha** etc.) ← PD ***muẓi/miẓi** ‘say, speak, utter’; ***muẓankk** ‘make noise, speak’ DEDR 4989; see **draviḍa** in **3.22(4)**. M53: MIA variations point to a foreign word or tribal name as the most plausible source, M86: unclear. (But see Witzel 1999b: 38.)

śava ‘corpse’ ŚB ← PD ***cav-a*** ‘corpse’ ← PD ***cā(v)*** ‘die’ DEDR 2426 (Burrow 1955). Note that the RV form with short **-a-** would be the expected form in Dravidian, with shortening of the root vowel before a derivative suffix. Thieme (1955) proposes derivation from the verb **śvayati** ‘swells’ (root śū), i.e. **śava** = ‘the swollen or swelling one’, and suggests that the Pali form **chava** ‘corpse’ is derived from an OIA form with the prefix **ud-**, e.g. OIA ***ud-śvayati*** → ***ucchvayati***. M53 and M86 accept Thieme’s suggestion. (This word did not appear in the list of Dravidian loans in OIA in later editions of Burrow 1955, possibly in response to Thieme’s proposal.) Though Thieme’s derivation is perfectly plausible, it is by no means compelling, and is more complicated than the Dravidian explanation. Thus the latter can be regarded as more likely, both semantically and phonologically. Witzel (1999b: 29) suggests an accidental resemblance between OIA and Dravidian, which of course cannot be ruled out.

(4) Epic and Classical Sanskrit

aṭavi ‘forest’ R ← PSD ***aṭar-vi*** ‘thicket’, ***aṭar*** ‘be close’ DEDR 84. M53: prob. Dravidian; M86: not sufficiently explained.

anala ‘fire’ (L. only, but cf. Pali **anala** ‘fire’) ← PSD1 ***a<u>n</u>al*** ‘fire, heat; burn, be hot’ DEDR 327 (Burrow 1946). Though the attestation in Dravidian is only at the PSD1 level, the word appears in early Tamil texts and shows native morphophonemic changes (e.g. Tamil past form **a<u>n</u>a<u>n</u>r**). Here Thieme (1955) provides another example of a plausible but unnecessary Indo-European derivation. M53: probably Dravidian; M86: problematic.

eḍa(ka) ‘a kind of sheep’ Katyśr; ‘sheep, goat’ MBh ← PD ****yāṭu*** ‘sheep, goat’ DEDR 5152 (Burrow 1946). Note that **ē-** is a normal development from PD ***yā-*** in some Dravidian languages. M53: prob. Dravidian; M86: poss. IE (cf. Greek **aíz**, Arm. **ayc** ‘goat’). Cf. **yādava** in **3.22(1)**.

kaṭhina ‘hard, firm’ Suśr (also **kaṭhora**, **kaṭhura**) ← PD ***kaṭṭ*** ‘harden’ DEDR 1148 (Burrow 1947a). M53: presumably Dravidian; M86: not explained.

kamala R, **kuvalaya** ‘lotus’ MBh ← SD1 ***kumaḷai/komaḷai*** ‘waterflower which closes by day’ ← PD ***kūmp/kōmp*** ‘close (as a flower)’ (Burrow 1947b). M53: prob. Dravidian; M86: allegedly Dravidian.

karīra ‘shoot of bamboo’ Suśr ← PD ***kaẓ(-al/il)*** ‘(bamboo) shoot’, cf. PD ***kuẓ*** ‘sprout’ DEDR 2149 (Burrow 1947a). M53: prob. Dravidian; M86: unclear, perhaps borrowed.

kalā ‘any practical art’ R ← PSD ***kal*** ‘learn’ DEDR 1297 (Burrow 1947b). M53: prob. Dravidian; M86: not explained. (The CD forms appear to come from ***kar*** or ***kār***, leaving some uncertainty about the earliest Dravidian form of this word.)

kānana ‘forest’ R ← PSD ****kān*(al)** ‘forest’ DEDR 1418, cf. PD ****kāṭu*** ‘forest, jungle’ (= uninhabited area) DEDR 1438, Burrow 1947a, M53, M86: prob. Dravidian.

kāla ‘black, dark-blue’ MBh ← PSD ***kāẓ** ‘black(ness), become black’ DEDR 1494 (Burrow 1946); cf. PD **kār*, **karV* ‘black’ DEDR 1278. M53: prob. Dravidian; M86: uncertain.

kuṇḍala ‘ring, earring’ āśvGṛ ← PD ***kuṇḍu** ‘round’ DEDR 1695 (Burrow 1947b). M53: prob. Dravidian; M86: unclear, perhaps foreign.

kuṭa ‘waterpot’ Yaśast ← PSD ***kuṭam** ‘pot, vessel’ DEDR 1651. M53: poss. Dravidian; M86: unclear.[10]

kuṭi ‘cottage, hut’ MBh ← PD ***kuṭi** ‘hut, house, etc.’ DEDR 1655 (Burrow 1946). M53: prob. Dravidian; M86: possibly Dravidian. (See **kuṭumb** in **3.22(3)**.)

koṭ(ṭ)a, koṭṭāra ‘fort, stronghold’ Kathās. ← PSD **kōṭṭay* ‘fort, castle’ DEDR 2207a (← PD ***kōṭṭay** ‘wall’ DEDR 2207b). M53; M86: prob. Dravidian. (See discussion of this word in **9.5C**.)

koṭara ‘hollow of a tree’ MBh ← PD ***koṭ(a)r** ‘hole in wall/tree/etc.’ DEDR 1660, cf. PD ***kuẓ** ‘cavity’ (Burrow 1947a). M53: Burrow’s suggestion worthy of attention; M86: prob. Dravidian. Cf. **kuṇḍa** in **3.22(1)**.

koṇa ‘corner’ MBh ← PD ***kōṇ** ‘angle, corner’ DEDR 2209 (Burrow 1946). M53: prob. Dravidian; M86: possibly Dravidian.

koraka ‘bud’ R ← PD ***kuẓ** ‘tender sprout/shoot/leaf/’, cf. PD ***kuẓ** ‘to sprout’ (Burrow 1946). M53, M86: prob. Dravidian.

candana ‘sandalwood’ Nirukta ← PSD ***cāntu** ‘daub, rub into paste; sandalwood (paste)’ DEDR 2448 (Burrow 1946; Emeneau 1971). Dravidian origin likely because of semantic development. M53, M86: prob. Dravidian.

capeṭā ‘slap with the open hand’ Pan ← PSD ***capp-aṭ** ‘clap, slap’ DEDR 2335, cf. PD ***capp-aṭ** ‘flat’ DEDR 2331 (Burrow 1947b). M53: poss. Dravidian; M86: foreign word?

cikkaṇa ‘unctuous, viscid’ MBh ← PD ***cikk-aṇ** ‘stickiness, gumminess, gum, birdlime, etc.’ DEDR 2488 (Burrow 1945). M53: poss. Dravidian; M86: not clear.

cumbati ‘kisses’ MBh ← PD ***cū(m)ppu** ‘to suck’ DEDR 2621a, PD ***cīppu** ‘to suck’ DEDR 2621b (Burrow 1945). M53: Dravidian connection unlikely; M86: prob. onomatopoetic.

tāḍaka, tāla, tālaka ‘lock, bolt’ BHS, Hariv ← PSD ***tāẓ** ‘bolt, lock’ DEDR 3179 (Burrow 1947a) M53, M86 (s.v. **tāla**): prob. Dravidian.

tāḍayati ‘beats, punishes’ MBh ← PD ***taṭ** ‘strike’ DEDR 3039 (Burrow 1947b, Emeneau 1971). M53: Dravidian most likely; M86: not satisfactorily explained. Cf. **taṇḍula** in **3.22A(2)**.

tāmarasa ‘lotus’ MBh ← PD ***tāmar** ‘lotus’ DEDR 3163 (Burrow 1946). M53: prob. Dravidian; M86: apparently Dravidian.

tulasī ‘the sacred basil plant’ BhP ← PD ***tuẓ**, **tuḷaci** ‘sacred basil, *Ocimum sanctum*’ DEDR 3357 (some forms possibly reborrowed from IA) (Burrow 1947b). Dravidian origin likely because of the large number of variant forms. M53, M86: prob. Dravidian.

tuvara/tubara ‘astringent’ Suśr ← PSD ***tuvar** ‘astringent’ DEDR 3352 (Burrow 1945). M53, M86: prob. Dravidian.

truṭati 'breaks' Pan (also ***truṇṭati**, CDIAL 6066) ← PD ***tu(ṇ)ṭ** 'break, come apart' DEDR 3310; cf. PD ***tuṇ** 'be broken; fragment' DEDR 3305. M53: prob. MIA or non-Aryan; M86: not explained.

draviḍā 'Dravidian' Mn; **dramiḍa**, **drāviḍa** MBh; Pali **damila**, **daviḷa**, **daviḍa** ← SD1 ***tamiẓ** 'Tamil language or people' (?← ***tammiẓ** 'own speech' ← PD ***tā̱n/ta̱n** 'one's own' DEDR 3196 + PD ***muẓ/miẓ** 'speech' DEDR 4989, cf. **mleccha** in **3.22A(3)**); see Southworth (1998). M53, M86: prob. connected with **Tamil**, otherwise unclear.

nirguṇḍī 'the shrub *Vitex nigundo*' Suśr ← PD ***nīr-kuṇṭ** 'water-hole' ← PD ***nīr** 'water' DEDR 3690 + PD ***kuṇṭam** 'cavity, pit' DEDR 1669 (see **kuṇḍa** in **3.22A(1))**. Cf. **nig(a)ḍe** as a place-name suffix in Maharashtra (OIA***nīr-guṇḍ** → OM ***nigguḍ(i)** → Marathi **nig(u)ḍ/ nigaḍ/ nigḍi**), possibly referring originally to a water hole or a place where *V. nigundo* was found (see **9.24C**). The plant usually grows near water. M53, M86: prob. connected with DEDR 3781 Ta. **nocci** etc. (Burrow 1947a).

paṭṭaṇa/paṭṭana 'town' ON, VarBrS ← PSD ***paṭṭi** 'cow-stall, habitation'; ***paṭṭiṇam/ paṭṭaṇam** 'town, city, seaport' DEDR 3868 (Burrow 1947a). M53 notes that the form **pattana** with dental **-tt-** is attested earlier (from the MBh on) and suggests that if the word is borrowed (Dravidian being the likeliest source), the earlier form with dentals represents phonological assimilation of the foreign word. M86: controversial. (See the discussion of 'spontaneous "fission" in **3.22A:** 'Phonological assumptions'.)

pallī 'small village, hamlet' Kathas ← PD ****paḷḷi*** 'hamlet, small village' ←***paḷḷi** 'house, sleeping place' ← ***paṭu** 'fall, lie' (Burrow 1946). M53, M86: most prob. Dravidian. See **9.21** for the use of this word as a place name suffix in Maharashtra.

puttikā 'the white ant or termite' Mn, **(piplikā)puṭa** '(ant)hill' ← PD ***pu̱ṟra** 'anthill' DEDR 4335 (Burrow 1945, Emeneau 1954). M53: poss. Dravidian; M86: Dravidian origin advocated by some.

malaya 'mountains bordering Malabar on the east' MBh ← PD ***mal** 'hill, mountain' DEDR 4742 (Emeneau 1954). M53: Dravidian.

maṣī/masi/masī 'ink' VarBrS, Lalit, Susr ← PD ***mas-i** 'ink, soot' DEDR 5101, cf. PD ***mā** 'black, dark' DEDR 4781 (Burrow 1946). M53: prob. connected with the Dravidian words; M86: Dravidian.

mālā 'wreath, garland' GrS ← PSD ***mālai** 'garland, necklace' DEDR 4827 (Burrow 1947a, Emeneau 1954). M53: connected with the Dravidian words; M86: prob. a foreign word.

mukuṭa 'crest, diadem' MBh ← PSD ***muk-aṭ** 'top (of building, hill), crown (of head)' DEDR 4888 (Burrow 1946). M53: prob. connected with the Dravidian words; M86: prob. Dravidian.

mukula 'bud' Suśr ← PSD ***muk-iẓ** 'bud; to bud' DEDR 4893 (Burrow 1946). M53: Dravidian; M86: prob. Dravidian.

muktā 'pearl' Mn ← PSD ****mutt*** 'pearl, tear, drop' DEDR 4959; cf. PSD ***mutt-u** 'pockmark, pimple, corn, etc.' DEDR 4961, PSD ***mutt-ai** 'lump' DEDR 4962

(Burrow 1946). M53: prob. from Dravidian, which may not be the original source; M86: hypercorrect form, attributed by some to Dravidian. See discussion in **3.22A:** ‘Phonological assumptions’ item (4).

muraja ‘drum, tambourine’ MBh ← PD ***mur** ‘to sound, speak’, cf. Ta **muracam/muracu/muravu/muruṭu** ‘drum’, Ka **more** ‘hum, buzz, sound, play lute’, Kol **moray** ‘produce musical sound’ DEDR 4973 (Burrow 1946, 1947b). M53: prob. Dravidian.

murangī ‘*Moringa pterygosperma*’ Suśr ← PD ***murun-kāy** ‘a tree or shrub with edible fruit’ DEDR 4982 (Burrow 1946). The second element is presumably PD ***kāy** ‘unripe fruit; to bear fruit’ DEDR 1459). M53: prob. Dravidian; M86: Dravidian.

valaya ‘bracelet’ MBh, ‘circle’ Kathas ← PSD ***vaḷay** ‘surround, enclose’, cf. Ta **vaḷai** ‘bangle, bracelet, circle’, Ko **vaḷ** ‘bangle’, Ka **baḷe** ‘ring, armlet, bracelet’ DEDR 5313 (Burrow 1946). M53: prob. MIA or non-Aryan, poss. Dravidian; M86: not convincingly explained.

śṛṅgavera ‘dried or fresh ginger’ Suśr (cf. Pali **singivera**) ← PD ***cinki-vēr** ‘ginger root’ ← PSD1 **cinki* ‘ginger’ (cf. Ta **inci** DEDR 429 ← ***cinki** (with normal loss of initial ***c-** and palatalization of **-k-** before **-i-**) + PD **vēr** ‘root’ DEDR 5535 (Burrow 1946). M53: **vera** is Dravidian, but the first element may be an east Asiatic cultural loan; M86: prob. connected to Dravidian. The OIA form is probably the result of folk-etymologizing, based on **śṛṅga** ‘horn’ (an Indo-European word, cf. Latin **cornu**, Eng. **horn**). PSD1 ***cinci-vēr** is presumably the source of Greek **zinziber**, E. **ginger**, etc.

hintāla ‘date-palm, *Phoenix paludosa*’ Hariv ← PD **cīnt(t)-tāẓ* ‘date-palm’ ← PD **cīnt(t)* ‘date’ DEDR 2617 + PD **tāẓ* ‘palm’ DEDR 3180 (Burrow 1945). The OIA **h-** may well represent a transitional stage between PD ***c** and SD1 **ø**. See **tāla** in **3.23A(5)**. While there is no problem, either phonological or chronological, in accepting the Dravidian word as the source of the OIA word, a question arises regarding the relationship between PD ***cīnt** and Munda forms like Santali **kindet’** ‘wild date-palm’ and Mundari **kindad’** ‘stemless palm’, which Kuiper considers to be the source of the Kurukh form **kīndā**.[11] Note also Bantu **mu-kindu** ‘date’ (Nurse 1983). These forms with initial **k-** clearly cannot be separated from the Dravidian forms with initial **c-**, but the relationship is not clear. M53: difficult; M86: definitely a foreign name.

heramba ‘buffalo’ Malatimadhava ← PSD ***erumai** ‘female buffalo’ DEDR 816, cf. PD ***eru-tu/ṭu** ‘bullock, ox’ DEDR 815 (Burrow 1947a). M53: poss. Dravidian; M86: uncertain.

3.22B. Early Dravidian borrowings from OIA

Listed here are probable OIA borrowings in early Dravidian, from DBIA, the DEDR, and the appendix to DEDR. These are listed as either Proto-Dravidian or

Proto-South Dravidian, depending on their distribution among the Dravidian languages. Words which are only attested in the literary Dravidian languages (in which Tulu is included for present purposes) are not included here, unless warranted for semantic or phonetic reasons. However, even for the cases listed here, it is not possible to show that they were necessarily borrowed into Dravidian at an early stage, since they might well have been borrowed by the individual languages from MIA or NIA sources at different times. In many cases, the route of the borrowed words was probably from Sanskrit into one of the literary Dravidian languages, and thence into the non-literary languages.[12] On the other hand, it may be noted that a number of these words are everyday terms referring to early technical innovations. See **8.24** for discussion of the implications of these words for prehistory; cf. also Witzel 1999b: 29. See the entries in DEDR and DBIA for further etymological information.

Proto-Dravidian

***accu** 'axle' (DBIA 7) ← OIA **akṣa** 'axle' RV
***aṇṭ** 'male' (of animals) (DEDR App. 7) ← OIA **aṇḍīra** (lex.) 'male' (see **8.4**)
***kuṭāri/kuṭāli** 'axe' (DEDR App. 32) ← OIA **kuṭhāra**
***cūci** 'needle' (DBIA 171) ← OIA **sūci** RV (← **sīv** 'sew')

Proto-South Dravidian

***arank** 'stage, platform, veranda' (DEDR App. 8) ← OIA **ranga**
***arac-an** 'king' (DEDR 201) ← OIA **rājan**
***argaḷ** 'bar, crossbar' (DEDR App. 9) ← OIA **argala**
***āṇi** 'nail' (DEDR App. 10) ← OIA **āṇi** (RV)
***kañc** 'bell-metal' (DBIA 67) ← OIA **kaṃśa** 'metal vessel' AV
***kaṭṭ** 'stick' (DBIA 68) ← OIA **kāṣṭha** 'piece of wood' ŚBr
***kump-at** 'gourd, pumpkin, *Cucurbita pepo*' (DEDR App. 28) ← OIA **kumbha-phala, kumbhaṇḍa** 'gourd'[13]
***paṇ-i** 'comb'(DEDR App. 49) ← IA (Pkt. **phaṇaga**, Marathi **phaṇi)**
***paṇṭi** 'cart, wagon' (DEDR App. 50) ← OIA **bhāṇḍa** 'goods, wares'[14]
***pōy** 'member of a spec. tribe/caste, headman' ← OIA **bhōgin** (DEDR App. 51)
***may-aṇ** 'wax' ← OIA **madana**, Pkt **mayaṇa**, Marathi **meṇ** (DEDR App. 52).

3.23. Other sources of lexical borrowing

The following words are attested early in both Indo-Aryan and Dravidian, without there being any possibility at present of determining whether they originated in OIA, Dravidian, or some third language – perhaps the "Indus" language(s). Only words attested in Classical Sanskrit or earlier are included here. The list includes a number of items from Southworth 1979b. (Entry items are OIA unless indicated otherwise.)

(1) Proto-Indo-Iranian (2 items)

khara, garda-bha 'donkey', Avestan **ḥaro;** PD ***kaẓ-ut-ay*** (DEDR 1364): according to Witzel (1999b:54), these words are from a Central Asian substrate (perhaps ultimately of west Asian origin) and "one of the few possible links of a Central Asian substrate with Dravidian (and with Vedic)". Note however the close resemblance between OIA **garda** and Central Dravidian forms like Naiki **gāṛdi**. Mayrhofer: (1) **khara**: M53: no conclusion; M86: Indo-Iranian word, otherwise unclear; foreign word? (2) **garda-bha**: M53: prob. Connected to **gard** 'cry out' (Thieme 1955); M86: Dravidian derivation of this early word doubtful ("wenig glaubhaft").

godhūma 'wheat' VS; cf. Avestan **gantumō**; Kannada **gōdi**, Brahui **k͟hōlum** (DBIA 123); cf. also Kd **kuli** 'paddy'(DEDR 1906). Witzel (1999b: 29, 55) shows a number of central Asian and west Asian (near Eastern) cognates, and suggests that the ultimate origin of this word is probably in one of these areas. M53 and M86 point to Iranian cognates and note the similarities to the Dravidian words, suggesting that the OIA word is a folk-etymological adaptation of a word from some (unidentified) foreign source.

(2) Rigvedic (6 items)

ulūkhala 'mortar' RV; SD1 ***ul-akk*** 'pestle' DEDR 672, cf. ***uram-kkal*** ← ***uram** 'mortar' DEDR 651 + ***kal*** 'stone' DEDR 1298; possibly both forms are from a third source, with PD folk-etymological restructuring along with OIA restructuring based on ***urū-khara** 'having a broad **khara**' (Thieme 1955); M53: prob. Dravidian; M86: problematic.

kūṭa 'mallet, hammer' RV ← PD ***koṭṭ** 'mallet' ← PD ***koṭṭ** 'beat' DEDR 2063; cf. PD ***kuṭṭ** 'strike' DEDR 1671; Santali and Mundari **kutam** 'to beat, hammer', Mundari, Ho **kutasi** 'hammer' (Burrow 1945). M53: origin unknown; Dravidian origin of this early word questionable; M86: unclear.

nīla 'dark, blue, blue-black; indigo' RV; PSD1 ***aṇile** 'ink-nut, ink-nut tree, *Terminalia chebula*' DEDR 119; probably from a pre-Aryan ***ṇīl** 'ink-nut tree'. M53, M86: not convincingly explained.

piśāci, piśāca 'demon' RV, AV: PD ***pēy, *pēn*** 'devil, god, spirit' DEDR 4438, 5529, 5530. M53, M86: not convincingly explained.

lāngala 'plough' RV; PD ***ñāñ-kVl*** 'plough' DEDR 2907, probably a folk-etymological formation: "earth-stone", based on PD ***ñān/ñāl** 'earth' DEDR 2907 + PD ***kal*** 'stone' DEDR 1298, cf. Ta **nānkuẓu**, Ma **ñāññūḷ** 'earth-worm' DEDR 2906). M53: prob. ultimately from Austro-Asiatic, cf. Khasi **lynkor**, Khmer **ankal/ānkal** etc.; M86: unclear; foreign word? See Witzel 1999b:38–9 for further links.

sindhu 'river, sea; the Indus River' RV, 'the country of Sindh' MBh; cf. Burushaski **sinda** 'river' (Pinnow 1953–54: 12ff.). M53 notes possible IE etymologies, but does not exclude a non-Aryan source; M86: not explained.

I have suggested (1990b: 228) a connection between OIA **sindhu** and the Dravidian words for the date (see **hintāla** in **3.22A(4)**).

(3) Other Saṃhita Vedic (6 items)

āmra ‘mango’ ŚB, ‘mango tree’ MBh, **mākanda** ‘mango tree’ Class: PD ***mām*** DEDR 4782, ***maṯ-kāy*** DEDR 3408. M53: connected with OIA **āmla** ‘bitter’; M86: not clear.

ikṣu ‘sugarcane’ AV: PD ***in(c)cu** ‘sweet juice’ ← PD ***in/en** ‘sweet’ DEDR 530 + PD ***cu** ‘juice’ (cf. ND ***cu-ra** ‘juice, sap’, Kur **currā** ‘vegetal sugar’ DEDR 2711, and PSD ***cu-na** ‘juice which exudes from plants’ DEDR 2717; note Te **incu** ‘sugarcane’ DEDR 530. OIA **ikṣu** may well be derived from Dravidian, though both might be derived from a third source, with folk-etymologizing of the Dravidian form. M53: unclear. M86: vowel variation in MIA and NIA implies an original ***ṛkṣu**, which would be unexplainable.

kanka ‘heron’ VS: PD ***korV-nk(k)*** ‘crane’ DEDR 2125, Burrow (1945). M53, M86: possibly onomatopoetic.

nīvāra ‘wild rice’ VS: PSD ***navarai/nivari** ‘rice’ DEDR 3614. M53: poss. Dravidian; M86: not clear.

vrīhi ‘rice’ AV: PD ***var-iñc*** ‘seed; rice’ (PD ***varici** ‘rice, grain’ DEDR 215, 5265; PD ***vanci/*manci** ‘seed’ DEDR 4639; cf. also PD ***val(ci)** ‘rice’ DEDR 5287). (For initial **v~m**, see Zvelebil 1970: 157; for loss of initial **v-**, see PSD ***(v)ari** ‘tax’, DEDR 216, 5266.) For the meaning ‘seed’, note Kur. **mānjī** ‘seed in general’ DEDR 4639, also Ta. **ēla-v-arici** ‘cardamom seed’ DEDR 907; the gloss of the Kur. form suggests that at the PD level the meaning may have been ‘seed’ (of unspecified grain) – and note Elamite **bar** ‘seed’. PD ***var-iñc*** is also linked to Malagasy **vari, vare** ‘rice’, Somali **barís** ‘rice’, and Ngaju–Dayak **bari** ‘boiled rice, food’. Persian **birinj** ‘rice’ (borrowed into Brahui as **brinj**) is presumably linked to PD ***var-iñc***. Greek **oryzon** presumably represents PD. ***varic** with the Greek neuter suffix **-on** (Southworth 1979b). See Witzel 1999b: 39–43 for further links. M53, M86: note Iranian cognates; possibly (eastern) culture word.

śarkarā ‘grit’ AV, ‘candied sugar’ Hariv: PD ***ceṯ-Vkk*** ‘sugarcane’ DEDR 2795. M86: difficult.

(4) Later Vedic (3 items)

arka ‘the plant *Calatropis gigantea*’ ŚB: PSD1 ***erukku** ‘id.’ (← ***cer-**?) DEDR 814. M53: prob. Dravidian; M86: similar to Dravidian words.

tuṇḍa ‘beak, trunk, snout’ TAr MBh: PD ***tuṇṭam** ‘snout, elephant’s trunk’ DEDR 3311, PD ***tuṭ** ‘mouth, beak, lip’ DEDR 3296; cf. PD ***cuṇṭ** ‘bill, beak, mouth’ DEDR 2664, PD ***cuṇṭ** ‘mouse, shrew (with long snout?)’ DEDR 2661 (Bloch 1925). M53: unclear, apparently non-Aryan; M86: not convincingly explained.

muñja ‘the grass *Saccharum sara* or munja’ SBr: SD1 ***muñci*** ‘reedy sugarcane, *Saccharum arundinaceum*’ DEDR 4916. M53: unexplained; M86: unclear; foreign word?

(5) Epic and Classical Sanskrit (30 items)

agasti ‘the tree *Agasti grandiflorum*’ Susr: SD1 **agace** DEDR 5 (Burrow BSOAS 12). M53: prob. Dravidian; M86: note similarity to Dravidian words.

elā ‘cardamom’ Susr: PSD ***ēla** DEDR 907. M53: unclear; M86: Dravidian.

kajjala ‘lampblack’ Susr: Ta **karical** ‘blackness’ ← PD ***kar** ‘black; be burned’ (Burrow 1946). M53: from Munda (Kuiper 1948: 28f.) or Dravidian; M86: prob. a foreign word.

karavīra ‘oleander, *Nerium odorum*’ MBh, **kaṇavīra** Pkt: PSD ***kaṇavīra** DEDR 1164 (Burrow 1947b). M53: prob. from a Dravidian source; M86: Dravidian.

kulattha ‘horsegram, *Dolichos uniflorus*’ MBh: PD ***koḷ** ‘id.’ (Burrow 1945). M53: prob. Dravidian; M86: connected to the Dravidian words. Cf. also Proto-Munda ***kodaXj** (Zide and Zide 1973). For the suffix cf. OIA **aśvattha** ‘*Ficus religiosa*’.

kustumbarī ‘coriander’ Suśr, **tumburu** ‘fruit of *Diospyros embryopteris*, coriander’ Pan: PD ***tump-i/ari/ili** etc. ‘*D. tomentosa, D. ebenum, D. melanoxylon*’ DEDR 3329. M53: unclear, prob. borrowed; M86: prob. an indigenous plant name.

ketaka ‘the tree *Pandanus odoratissimus*’ MBh ← PSD ***kaita(-ke)** ‘fragrant screwpine, *P. odoratissimus*’ DEDR 2026 (Burrow 1946, Emeneau 1954). M53: prob. Dravidian; M86: Dravidian.

kuṣmāṇḍa ‘a kind of gourd or pumpkin’ MBh: PD ***kumpa-ḷa/ṇṭa/ṭṭa** ‘gourd, melon’ (DEDR app. 28). M53: prob. Austro-Asiatic.

goṇī BHS ‘sack’: PSD ***kōṇi** ‘sack, sackcloth’ (DEDR App. 33). M53: prob. Dravidian; M86: not satisfactorily explained.

ciñcā, cintidī ‘the tree *Tamarindus indica*’ Bhpr: PD ***cin-tta*** DEDR 2529. M53: OIA and Dravidian words from a Munda source; M86: foreign word.

tāmarasa ‘red lotus’ MBh: PD ***tāmar** ‘lotus’ DEDR 3163 (Burrow 1946). M53: prob. Dravidian; M86: apparently Dravidian.

tāla ‘palmyra or toddy palm, *Borassus flabelliformis*’ Mn: PD ***tāẓ*** ‘toddy palm, *B. flabelliformis*’ DEDR 3180 (Burrow 1945; some forms possibly borrowed from IA); cf. SD1 ***tāẓai** ‘coconut palm’ DEDR 3183. M53, M86: perhaps linked with the Dravidian words. See **hintāla** in **3.22A(4)**.

tuṇḍikeri Susr, **tuṇḍikā** Lex ‘the gourd *Momordica monadelpha*’: PD ***toṇṭa** DEDR 3499. M53: poss. Dravidian; M86: prob. Dravidian.

nakra ‘crocodile’ Mn MBh: PSD ***nekaẓ** ‘marine animal which attacks swimmers’ (Burrow 1946). M53: prob. Dravidian; M86: not explained.

nārikera/nālikela Susr, **nārikela** MBh, **nāḍikerī/nālīkera** BHS ‘coconut palm and fruit’: PD ***nār(i)** ‘fiber or rope from coconut palm or other plant’ DEDR 3651 + Tamil **kēḷi** ‘coconut palm’. M53, M86: indigenous word.

nimba ‘the tree *Azadirachta indica*’ Gobh: PSD ***vē-mpu*** ‘neem, *A. indica*’ DEDR 5531 (Burrow 1947b). M53, M86: unclear.

paṭola ‘the gourd *Tricosanthes dioeca*; its fruit’ Susr: PD ***puṭala/poṭala*** DEDR 4250. M53, M86: prob. Dravidian.

panasa ‘the breadfruit tree, *Artocarpus integrifolia*’ MBh, ‘its fruit’ Susr: PD ***pal-ac/pan-ac*** DEDR 3988. M53: poss. Dravidian, ultimately Austro-Asiatic; M86: prob. Dravidian.

punkha ‘feathered part of arrow’ MBh: SD1 ***puẓuku** DEDR 4314 (Burrow 1945). M53: Burrow’s derivation worthy of note; M86: not satisfactorily explained.

pūga ‘*Areca catechu*, its nut’ Susr: PD ***pōka** DEDR 4048 (Burrow 1947b). M53: poss. Dravidian; M86: prob. Dravidian.

baka ‘crane’ Mn ← PSD ***vakk-ā/u** ‘stork/crane’ DEDR 5206 (Burrow 1947a). M53, M86: prob. onomatopoetic.

bakula, vakula ‘the tree *Mimusops elengi* or its flower’ MBh (also **makula** L: H **maulsarī**: PSD1 ***pogaḍa** DEDR 4453, ***mukul, *makiẓ** DEDR 4619 (Burrow 1947b). M53, M86: possibly linked to the Dravidian words.

mallikā ‘*Jasminum zambac*’ MBh ← PSD ***mullai** ‘jasmine’ DEDR 4987 (Burrow 1946). Cf. also PSD ***vaḷḷi** ‘creeper’ below. M53: poss. connection with Dravidian; M86: prob. Dravidian.

yavanāla ‘the grain *Andropogon bicolor*’ Susr, **yonala** Lex: PD ***co<u>nn</u>a-l*** ‘millet, maize’ (← ***cōḷ-nel**?) DEDR 2896; note Ta. **cōḷam, connal**, Te **jonna, jonnalu**, and Brahui **cōṇḍ** ‘lucerne (alfalfa)’ (cited in DED 2359 but omitted from DEDR 2896); cf. PD ***nel** ‘paddy, rice in the husk’ DEDR 3753). M53: ← OIA **yavana** ‘Greek, Ionian’; M86: indigenous plant name?. Derivation from Dravidian and subsequent folk etymologizing in OIA is more likely, given the structure of the word in Dravidian – though the ultimate origin might lie elsewhere.

laśuna Gaut, **rasona** Susr ‘garlic’: ?PD ***lacuṇi** (Kui **lesuṛi**, Mlt **nasnu** Burrow (1946), not in DEDR). M53: apparently a culture word; M86: not adequately explained.

vallī ‘creeper’ Mn: PSD ***vaḷḷi** ‘creeper, climbing plant’ DEDR 5316 (Burrow 1946). M53: MIA or non-Aryan; poss. Dravidian. See **mallikā** above.

varuka ‘a species of inferior grain’ Susr (MW): PSD ***varaku** ‘millet’ DEDR 5260 (Burrow BSOAS 12). M53: prob. connected with Drav; M86: prob. Dravidian.

vīraṇa/viraṇa ‘a fragrant grass, *Andropogon muricatus* MBh (MW): PSD ***viẓ-al** DEDR 5428. M53: prob. connected with Dravidian; M86: unclear; foreign word?

śṛṅgavera ‘dried or fresh ginger’ – PD ***cinki-vēr** ‘ginger root’: see **3.22A**(**4**). This is a compound word in which the second element is probably Dravidian, while the first part is of uncertain origin.

sarja ‘the tree *Vatica robusta*’ MBh: SD1 ***āccā/ārc** (? ← ***cārcā**) ‘sal tree, *Shorea robusta’* DEDR 343 (Burrow BSOAS 12). M53, M86: unclear.

3.24. Sources of the OIA lexicon

A useful approach to the study of the sources of the OIA lexicon would be to examine specific domains of vocabulary to discover the proportions of items coming from different sources at different time periods. A comprehensive study of OIA plant names, for example, could be immensely informative in this respect. As an indication of what might be done in this connection, the agricultural materials collected by Masica (1979) for Hindi (restricted to those terms which are attested in OIA) have been analyzed by period of attestation and origin; the results of this analysis are presented in the following table. Note that in all periods, the numbers of unknown words are at least as high as those of presumed Munda or Dravidian origin. These data show that one-third of the OIA names for plants in this list are of unknown origin, and that only 40 percent have known Indo-European etymologies. This is of course not surprising, given that many of the plants in this list were unknown to earlier stages of Indo-European.

Sources of OIA agricultural vocabulary (based on Masica 1979)

Period	*Source*					
	IE/IIr	*Drav.*	*Munda*	*Other*	*Unknown*	*Total*
Proto-IIr	4	—	—	—	—	**4**
Rigvedic	16	5	2	—	5	**28**
Other Samhita Vedic	7	2	1	—	8	**18**
Late Vedic	4	—	1	—	3	**8**
Epic and Classical	10	6	3	2	6	**27**
Later Sanskrit	8	3	6	—	19	**36**
Totals	**49**	**16**	**13**	**2**	**41**	**121**
	(40%	13%	11%	2%	34%	100%)

3.3. Structural evidence

3.31. Phonological convergence

On the phonological side, the most striking feature of Indo-Aryan is the retroflex–dental contrast, which exists from the earliest Vedic – though the frequency of retroflex consonants increases with time from the earliest texts up to the classical period – and which eventually comes to be expressed in stops, nasals, laterals, and sibilants. This distinction sets the Indo-Aryan languages apart from all other Indo-European languages – including the closely-related Iranian languages, which together with Indo-Aryan form the Indo-Iranian subfamily of Indo-European. Generations of scholars have thought it more than a coincidence that this distinction is found in precisely the same region as the Dravidian languages, for which it must be considered a proto-feature[15] – though it is

possible that at least some Dravidian retroflex and alveolar stops derived from combinations like **r** + **t** or **l** + **t** (Tikkanen 1988). Emeneau's view of the situation was the following:

> ... it is beyond doubt that, even where Indo-European material yields Sanskrit retroflexes, pre-Indo-Aryan and pre-Dravidian bilingualism provided the conditions which allowed pre-Indo-Aryan allophones to be redistributed as retroflex phonemes.
>
> (Emeneau 1956: 6, reprinted from M. B. Emeneau, 'India as a linguistic area', *Language* **32**: 3–16, with permission of the Linguistic Society of America)

The main argument against this claim has been that retroflex consonants could come about by purely regular and "normal" processes of change, and a number of scholars believed that they had done so in this case. It was to answer this point that Kuiper (1967) attempted to show that the general thesis of a Dravidian substratum could be demonstrated by the evidence of certain syntactic changes in OIA. Subsequently, Hans Hock (1975) challenged the whole substratum theory, maintaining that the evidence for Dravidian phonological and syntactic influence on OIA was inadequate, and that (at least for the Rigvedic period) even the evidence of Dravidian-OIA lexical borrowing was insufficient.

Munda languages also have a dental–retroflex distinction (see Pinnow 1959), though it is apparently not original except for **ḍ** (Zide 1969: 414, 423). Witzel (1999b) notes that retroflex consonants occur in the oldest books of the Rigveda in words where there is no phonological conditioning. Since there are, as Witzel claims, no Dravidian loanwords in those early books, Dravidian languages cannot be the source of the distinction. From what has been said in **3.21**, it seems likely that Munda/AA languages had greater influence on early OIA than Dravidian. Furthermore, (Para-)Munda itself probably borrowed words from earlier languages, and may have been structurally influenced by them also. If, however, the retroflex–dental distinction in Indo-Aryan came about as a result of contact with Munda languages, then these languages must have received the distinction from an earlier language, since it was not original in Munda. Witzel concludes that "the people of the (northern) Indus civilization must have spoken with retroflexes." Following this argument, we would infer that the "Indus" language(s) had the distinction.

Hock and others have argued that the sequence of changes which led to the development of the dental–retroflex contrast in OIA can be considered internally motivated, that is, that this sequence involves nothing other than normal conditioned sound change, whose working can be observed in words of Indo-European origin (Hock 1975: 99ff.). While this may well be the case, the possibility of external influence on this development is not ruled out.[16] We must look to the historical context of the change for further evidence. The fact that the frequency of retroflex consonants increases throughout the history of Indo-Aryan suggests that

what we see in the Rigveda is the leading edge of a change. Table 3.1 shows that this is the first of several changes which took place in Indo-Aryan in a similar way – that is, increasing in frequency and/or in regularity at successive historical stages – all bringing the phonology of Indo-Aryan closer to that of Dravidian. Change 2 in the table involves the assimilation of dissimilar consonant clusters: For example, OIA **sapta** 'seven' → Pali Pkt. **satta** → Panjabi **satt**, Hindi–Urdu **sāt**; OIA **miśra-** 'mixed' → Pali Pkt. **missa**.

Change 3 involves the voicing of intervocalic voiceless stops along with the spirantization or loss of voiced stops, as in OIA **śuka** 'parrot' → Pali **suka/suva**, Pkt. **su(g)a**; OIA **nija** 'one's own' → Pali **nija/niya**, Pkt. **ia**; OIA **ruta** 'voice' → Pali **ruta/ruda**, Pkt. **rua**. (Even the original voiceless stops lose their closure by the time of NIA: Hindi **kiyā**, Marathi **ke**(lā) 'done'← OIA **kṛta,** Pa Pk **kata/kita)**. Change 4 is the deaspiration of aspirated consonants as in OIA ***kathika** 'agreement' → Pkt **kat(h)ika**, Pali **hattha** 'hand' (from OIA **hasta**) → Hindi–Urdu **hāth**, Ass. Bg. Or. M. Ko. **hāt**. In the modern Indo-Aryan languages of the "outer" group, the deaspiration of non-initial stops is very frequent, and in some cases regular: see **5.22(5)**. The last change listed in the figure involves the shift of the stress accent in a word to the initial syllable: see **5.22(3)**. While any single one of these changes means nothing, and even combinations of two or three of them may be found in other language families (e.g. changes similar to 2 and 3 are found in the Romance languages), the co-occurrence of five similar changes

Table 3.1 Chronological development of Indo-Aryan phonological features attributed to Dravidian or other pre-Indo-Aryan languages

	Vedic (1500 BCE–)	*Pali (500 BCE–)*	*Prakrit (300 BCE–)*	*"Outer" IA*	*Dravidian (3000 BCE)*
Retroflex–dental contrast	- - - >	- - - - > (increasing in frequency over time)	- - - - >	- - - - >	Regular Proto-Dravidian feature
Assimilation of stops	Sporadic	Regular	Regular	Regular	Regular Proto-Dravidian feature
"Weakening" of intervocalic stops	Absent	Sporadic	Regular	Regular	Sporadic in late Proto-Dravidian; regular in modern Dravidian
Deaspiration of stops	Absent	Sporadic	Sporadic	Regular in some positions	No aspiration in pre-contact Dravidian
Word-initial stress	(variable musical accent)	MIA accent (see text)	MIA accent	Initial stress coexisting with MIA accent	Regular Proto-Dravidian feature

in two neighboring language groups is more than one could expect as a result of chance. If on the other hand, the occurrence of retroflex consonants in early OIA cannot be attributed to Dravidian, then it is possible that some or all of the features listed in Table 3.1 are to be traced back to the "Indus" language(s). Hock (1993) has suggested that the development of retroflex (as well as alveolar) consonants may have been a joint development involving both Dravidian and Indo-Aryan, along with eastern Iranian languages.[17] See **3.4** for further discussion of this possibility.

3.32. Syntactic influence (for theoretical background see 4.26)

Emeneau has discussed a number of cases in which the grammatical structure of OIA seems to have been influenced by Dravidian (see especially Dil 1980), and Kuiper has argued for two cases of Dravidian structural influence: (1) the use of the quotative particle **iti** (Avestan **uiti**), and in particular the order of the elements accompanying it, and (2) the absolutive participles or gerunds. Hans Hock has argued against these (and other) cases by showing that all the developments in OIA claimed to result from Dravidian influence have parallels in other ancient Indo-European languages.

Table 3.2 depicts the stages of the so-called "quotative" construction, which appears in sentences reporting direct discourse (DD) with a verb of saying (SAY), with or without an intervening particle. It was noted early in Indo-Iranian studies that Avestan, the oldest form of Iranian, and OIA share a quotative particle, **iti/uiti**, of unknown origin, and that Avestan uses **uiti** *before* the direct discourse and *after* the verb of saying, whereas OIA has not only that order but also the order DD + **iti** + SAY, which happens to be the order of a similar construction in Dravidian. (In Tamil, the element **en̲r̲-u** can be translated literally as "having-said" or "unquote." This form is a "conjunctive participle," a verbal form which is frequently used in South Asian languages to conjoin simple sentences into complex sentences. Again, Munda languages have a similar construction (see Dalai 1998), so Dravidian is not the only possible source. On the other hand, there exists a possible Dravidian source for the particle **iti/uiti**, which is otherwise

Table 3.2 Syntactic convergence: the quotative construction

Proto-Indo-European	*Pattern A* *DD + SAY* *[earlier?]*	*Pattern B* *SAY(+ ptcl) + DD* *[later?]*
Avestan	[absent]	SAY(+ **uiti**) + DD
Vedic	DD + **iti** + SAY	SAY(+ **iti**) + DD
Later Sanskrit	DD + **iti** + SAY	[absent]
Dravidian	DD + **eNR-u** + SAY	[absent]

without an etymology.[18] Thus it may be that in this case, Proto-Indo-Iranian borrowed not only the syntactic rule of quotation marking from Dravidian, but also the actual surface form which marks the end of the quote. The construction itself may not be original in Dravidian, but may be another case of the influence of the pre-Dravidian "Indus" language(s).

From this state of affairs Kuiper concluded that the Avestan order was the only order used in the Indo-Iranian parent language, and that OIA innovated the other order under Dravidian influence. Hock (1975: 107 ff.) pointed out that *both* orders are found in other ancient Indo-European languages; thus it is possible that Proto-Indo-Iranian inherited both orders, and Avestan lost one of them independently of any contact with other languages. From the viewpoint of Indo-Iranian, then, it makes sense to posit Pattern B as the only order, and to account for Pattern A in OIA as a Dravidian-inspired innovation; whereas according to Hock, if we look at it from a proto-IE perspective, it would make sense to assume both patterns in Proto-Indo-Iranian, and accept the loss of Pattern A in Iranian.

Again, possibly the historical context may provide clarification: in the subsequent history of OIA Pattern B, the order found in Avestan, almost completely disappeared. Thus here again we seem to be looking at the leading edge of a change in the earliest Indo-Aryan. The same seems to be true of the cases discussed by Emeneau. Thus while it is possible to argue, as Hock does, that there are earlier parallels in older IE languages, when we look at these linguistic data in the overall context it is difficult to dismiss the role of indigenous (Dravidian or other) influence. Witzel (1999a,b) points out that the absolutives are not found in early Iranian, though they occur in both Dravidian and Munda; the quotative construction also occurs in some Munda languages (Masica 1976: 189). Here again, as in the case of the retroflex consonants (**3.31**), we may be dealing with the influence of the "Indus" language(s).

3.4. Conclusions

All authorities acknowledge that the earliest Indo-Aryan language contained lexical elements derived from other languages. For those items referring specifically to South Asian flora, fauna, or peoples, the probability is that these names entered OIA (or in some cases, proto-Indo-Iranian) from languages spoken by groups which preceded the arrival of Indo-Iranian speakers into the subcontinent and its borderlands. Languages connected with Munda or Austro-Asiatic appear to be the earliest identifiable sources of foreign words in OIA, though some of these words may have originated in other unidentified languages (see **2.21**, Witzel 1999b: 10–19, 54–9). Words from (Para-)Munda continued to appear in texts from the earliest period of OIA down to the late Vedic period, indicating that this contact took place successively in the areas of present-day West and East Panjab, Sindh, Haryana, and Uttar Pradesh.

Clear evidence of contact with Dravidian languages begins in the middle Rigvedic period (*c.*1200 BCE), in contexts which suggest that Sindh may be the

first source area for this material (**3.22**, Witzel 1999b: 21ff.). Dravidian loans in OIA include terms related to pastoralism, agriculture, and indigenous plants. It is likely that many of the Dravidian loanwords found in OIA were mediated by speakers of "Outer" Indo-Aryan (see Chapters 5 and 6, and note the word **yādava** in **3.22A(4)**). Dravidian speakers also borrowed words from OIA, including the words for 'axle'and 'lynchpin' (**3.22B**), possibly as early as the Proto-Dravidian period. Like the (Para-)Munda words in OIA, Dravidian words also continue to appear in OIA texts through the period of Classical Sanskrit.

There is a possibility that there was earlier contact between Dravidian and OIA, which would have been separate from that of the middle Rigveda, since the earliest books of the RV do not contain recognizable Dravidian borrowings according to Witzel (1999b: 21ff.). The words listed in the Appendix to this chapter point to the possibility of contact with Proto-Indo-Iranian, before the separation of Proto-Iranian from Proto-IA. Such contact might have taken place in an area like Badakhshan, one of the probable sources of mineral products for the Indus Valley cities (Possehl 1999: 236; Salah *et al*. 1977: 281). We may also bear in mind the possibility of a connection between Dravidian and Uralic (**2.4**). The list of these words is not long, and some are controversial; thus some scholars will probably dismiss this lexical evidence as the result of coincidence. Nevertheless, it is worth mentioning, as the possibility of such contact cannot be ruled out by our present knowledge.[19]

A large number of the foreign words in OIA are not demonstrably from either Munda/AA or Dravidian sources. This category includes many words for indigenous South Asian flora. The sources of these words cannot be identified as yet, and may never be identified. An examination of agricultural vocabulary in modern Hindi indicates that, even among words which existed in OIA, approximately 34 percent cannot be traced to known languages (**3.24**). The most we can say about these words is that they are from pre-OIA and pre-Dravidian languages which were located in the general areas in which speakers of OIA found themselves at the time. Some of these words may have been mediated by groups of Indo-Aryan speakers other than those associated with the OIA texts, for example speakers of "outer" Indo-Aryan in Sindh, Gujarat, Maharashtra, Orissa, and Bengal. Again, these words are found at all chronological levels of OIA from the Vedas to Classical Sanskrit.

Changes in the phonological (**3.31**) and grammatical (**3.32**) structure of Indo-Aryan languages have been thought to show the influence of other languages, particularly Dravidian, though there is no general agreement on this question. Regarding the origin of retroflex consonants, it now seems most probable that they existed in the hypothetical pre-Indo-Aryan (and possibly pre-Dravidian) "Indus" language(s), and that both Indo-Aryan and Dravidian languages acquired them from contact with these languages (see **3.31**). If so, Hock's suggestion (see **3.31** end) that the development of retroflex consonants in OIA was part of a joint Indo-Aryan–Dravidian phonological change becomes more plausible.[20] This notion, along with the conclusion drawn in **3.32** regarding the sources of certain grammatical constructions, leads to the possibility that the South Asian "linguistic

area" or linguistic interaction zone existed even before the entry of Indo-Aryan languages into South Asia. Within the subcontinent, it involved speakers of (Para-)Munda, Dravidian, and 'Indus' languages, as well as the ancestors of Burushaski and other linguistic isolates mentioned in **3.2**, and was probably linked to speech communities of Central Asia.[21]

3.5. Summary

This chapter is a discussion of linguistic evidence found in Old Indo-Aryan texts which indicate contact between speakers of OIA and other languages. The chapter's conclusions are summarized graphically in the map (Figure 3.1). Note that this map covers a period of several millennia, since it includes the probable earliest locations of Munda/AA and Dravidian languages in the subcontinent, as well as the inferred locations of earlier languages such as the "Indus" language(s), along with the modern locations of Dravidian, Munda, and Tibeto-Burman languages, and isolated languages such as Nahali, which are not necessarily the same as their ancient locations.

Section **3.2** takes up the lexical evidence, looking first of all at loanwords in OIA which seem to be from Munda or Austro-Asiatic (AA) languages (**3.21**). These are the earliest identifiable foreign words in OIA, appearing in the oldest books of the Rigveda. Witzel and Kuiper have given the name "Para-Munda" to the language(s) which served as source(s) for this material. Though many of these words do not have specific Munda/AA etymologies, the hypothesis of AA origin is supported by the presence of prefixes of types found in Munda and other AA languages, which are not found in the other known language families of the area. These words appear in OIA texts belonging to the entire Vedic period, indicating the presence of Munda/AA speakers in all the regions associated with Vedic texts, from Panjab to eastern Uttar Pradesh – as well as further to the east, given the connection of Munda with the rest of Austro-Asiatic.

Dravidian loanwords in OIA appear at a somewhat later date, from the middle Rigvedic period (about 1200 BCE), in contexts which suggest a more southerly location, possibly Sindh (**3.22**). These words also continue to appear in OIA throughout the Vedic period and into the Epic and Classical Sanskrit periods. A small group of controversial words suggest the possibility of an even earlier contact between Dravidian and OIA, in the period of Proto-Indo-Iranian, which if it occurred must have been separate from that reflected in the Rigveda.

Apart from words attributable to languages of limited extent such as Burushaski and various (named or unnamed) Tibeto-Burman languages, an additional body of foreign words found in all periods of OIA are of unidentified origin, probably pre-Indo-Aryan and pre-Dravidian, and in some cases perhaps pre-Munda/AA (**3.23**). An examination of agricultural vocabulary in modern Hindi indicates that, even among words which existed in OIA, approximately 30 percent cannot be traced to known languages (**3.24**). Thus it is highly likely that a number of languages existed in South Asia before the arrival of Indo-Aryan

and Dravidian speakers. Only a few of these languages, such as Burushaski and Nahali, can be identified by name. The name "Indus" is used here to designate this group of languages.

Changes in the grammatical structure of Indo-Aryan languages have been thought to show the influence of other languages, particularly Dravidian, though some scholars disagree (**3.3**). Two such changes are discussed here: the development of the dental–retroflex contrast in phonology (**3.31**) and the so-called "quotative" construction (**3.32**). In both these cases, it is suggested that the history of these features in Indo-Aryan indicates that they are changes which began soon after the entry of Indo-Aryan speakers into the South Asian subcontinent, and thus are likely to have been triggered by local languages. Overall, the lexical and structural evidence of OIA and the Dravidian languages suggests that a linguistic interaction zone or "linguistic area" existed in South Asia before the arrival of Indo-Aryan languages in South Asia. Within the subcontinent, it involved speakers of (Para-)Munda, Dravidian, and "Indus" languages, as well as the ancestors of Burushaski and other linguistic isolates mentioned in **3.2**, and was probably linked to speech communities of Central Asia.

Appendix: possible Dravidian borrowings in Proto-Indo-Iranian

The evidence for contact between the Proto-Dravidian and Proto-Indo-Iranian speech communities is not very solid; more evidence would be required to prove the case. And yet, the following examples are difficult to dismiss outright, however implausible the hypothesis may be. An earlier, larger list has been pared down to these seven items, for which a good case can be made for Dravidian origin – even where other sources have been suggested. The first two items are what are referred to as 'function words' in Chapter 4, which suggest a possible contradiction, since the borrowing of function words usually implies the borrowing of other kinds of words, such as substantives and other content words (see **4.23**). Only a few of this latter type of word appear here, though of course there may be others for which adequate proof is lacking (see e.g. the words for 'donkey' in **3.23(1)** and Proto-Dravidian ***kaẓutai** in **8.21A1**); there are also other words which suggest early contact between the two speech communities, even though the ultimate source may be elsewhere (see **3.23**).

To accept this evidence, one would need to assume that there was a reason for the paucity of other kinds of borrowings. The most likely reason is that these items were not recognized as borrowings, and therefore were not considered inappropriate for ritual texts (see **4.33**). This entails an additional assumption, however: that the relations between the Proto-Dravidian and Proto-Indo-Iranian speech communities were close enough to permit the borrowing of function words – implying, as pointed out in **4.29**, some sort of economic interdependency, such as the relationship that might exist between pastoralists and agriculturalists. However difficult it might be to accept these assumptions, the following list may

be sufficient evidence at least to keep alive the hypothesis of a link between Dravidian languages and other languages outside of South Asia (see **2.4**):

(1) **iti** 'thus', quotative particle RV; Av. **uiti.** Cf. Classical Tamil **vi<u>tt</u>i** 'having caused to hear' (causative of the verb **vi<u>n</u>** 'hear' ← PD ***vi<u>n</u>** (DEDR 5516). See further discussion in **3.32**.

(2) **tan** 'self, own', **tanu** 'body, person, self' RV; also used as a reflexive, a function later gradually taken over by **ātman-**, but retained in the northwestern cognates of **tanū-** (e.g. Torwali **tanu** 'one's own'); Av. **tanū** 'body, person, self' (also used reflexively) ← PD ***tā<u>n</u>/ta<u>n</u>** '(one)self' DEDR 3196 (cf. PD **tām/tamm** 'themselves' DEDR 3162), Southworth 1979b: 198. M53 doubts the traditional connection with **tanu** 'thin, small, slender' (Lat. **tenuis**, etc.). The Dravidian word may go back to proto-Elamo-Dravidian: cf. Elamite **(u) atta-ta** 'my father' (Elamite **u** = nom. 'I' or gen. 'my', **atta** 'father', Hallock 1962). There is no question about the antiquity of this word in Dravidian. Since there is no sound Indo-European etymology, Dravidian origin is the most likely possibility. (One might argue that an Indo-Iranian word for 'body' came to be used as a reflexive pronoun under the influence of Dravidian **tā<u>n</u>/ta<u>n</u>**, but even such a possibility would indicate contact between Dravidian and proto-Indo-Iranian.)

(3) **maṭaci-** 'locust' ChUp., Av. **madaha** ← PD ***miṭuci** 'locust' DEDR 4850b, PD ***miṭi-** 'leap' DEDR 4850a, Burrow (1947a). Dravidian origin is most likely because of the derivation. M53: accidental resemblance, Dravidian origin, or Wanderwort.

(4) **māyā-** 'supernatural power, skill' RV ← PD ***mac/may-** 'mystify, confuse, intoxicate, delude' DEDR 4706; cf. also PD ***māy-** 'disappear, be lost' DEDR 4814. M53: origin uncertain.

(5) **mayūra-** 'peacock' RV ← ***mañil/mayil** 'peacock' DEDR 4642 ← ****ñam-Vl***. This connection was proposed in Burrow (1946), and subsequently rejected by Thieme, who claimed that although the word is likely to be a loan in OIA, it cannot be proven to be from Dravidian (1955: 440). Hock, citing Bailey (1957), notes the existence of a Saka **mur-āsa-** 'peacock', suggesting that "the word for 'peacock' was borrowed into Indo-Iranian prior to the Indo-Aryan arrival in India and that the Dravidian cognates are in fact borrowings from Indo-Aryan" (1975: 86). M53 points out that, while the connection with Dravidian cannot be ignored, there are also Austro-Asiatic words of similar shape, for example, Mon **mrā** (cf. OIA **marūka-**Lex.). These later discussions ignore the details of Burrow's original proposal, which shows that in addition to a PD ***mañil** 'peacock' (DEDR 4642) there is also a form ***ñamil** 'peacock' (DEDR 2902). Furthermore, at least two other Dravidian words show a similar phenomenon: PSD1 ***mayir (← *mañir?)** 'hair' DEDR 4707, PSD1 ***namir (← *ñamir?)** 'hair' DEDR 3615, and Ta. **miñiru/ñimiru** 'beetle, bee' DEDR 4843. Burrow cites Uralian cognates (Samoyed **ñīberu**, Ostyak-Samoyed **ñimere**) for the second case in support of the position that

the **ñ-** is original, and also provides another Uralian–Dravidian case of the same type: PD ***muñal** 'hare' (Ta. **muyal**) DEDR 4968 compared with Lappish-Norwegian **njoammel** 'hare', Mordwin **numolo** 'hare', etc. (1946: 608ff.). The change of intervocalic **-ñ-** to **-y-** occurs regularly or sporadically in a number of Dravidian languages (see Zvelebil 1970:137); thus the form **may(-il)** ← ***mañ-** ← ***ñam-** 'peacock' is clearly of Dravidian origin. This solution perhaps begs the question of the relationship between Uralic and Dravidian, which has been proposed by Tyler (1968). Whether such Dravidian–Uralic resemblances point to an earlier genetic relation or earlier interethnic contact (such as that between Uralic and Proto-Indo-Iranian, see Burrow 1973b) cannot be determined on presently available evidence, though some sort of early contact seems probable.

A point not mentioned by Burrow is that the earliest form in this sequence, ***ñam-**, is relatable to a PSD ***ñem-ir-** (← ***ñam-** with normal vowel change after initial **ñ-**) 'raise, straighten, extend, unfold' DEDR 2922, which describes the male peacock's tail display behavior. The **-l-** suffix of PD ***ñamil** and ***mañil** is clearly separable since some of the cognates lack it (e.g. Te. **nammi** DEDR 2902). And though none of the Dravidian words for 'peacock' has a suffix in **-ur-** or **-ūr-** (?Tu. **mairu** DEDR 4641), an **-ur-** suffix does occur in other Dravidian names for fauna, e.g. Ka. **munguri** 'mongoose' DEDR 4900 (in addition to **mungi, mungisi, munguli, mungili, mungi, mungali**). Alternatively, as Burrow suggests, the **-r-** of OIA **mayūra-** may be simply the expected Rigvedic equivalent of a non-IA **-l-** (see **6.14B**).

Thus the derivation of OIA **mayūra-** 'peacock' from Dravidian satisfies both phonological and semantic criteria better than any other proposal. In fact, the words for 'peacock' also have the meaning 'blue-green color' in several Dravidian languages. This seems to be parallel to other color words in Dravidian, such as DEDR 4635 Ta. **mañcaḷ** 'turmeric; yellow color'. On the other hand, Burrow is presumably correct in connecting the OIA **marūka-** 'peacock' to Austro-Asiatic (cf. Santali **marak**).

(6) **mukha-** 'mouth, face', Par. **muh** ← PD ***mun-kk-** 'nose' DEDR 5024, cf. PD ***mun-** 'in front' DEDR 5020a (Burrow 1947a). Though there are no Indo-European cognates outside of Indo-Iranian, M53 considers Dravidian origin unlikely in light of Vedic attestation, especially with Iranian cognates. Nevertheless, Dravidian origin is the most likely source in view of the age of the Dravidian word and its derivation within Dravidian. Note similar words in Austro-Asiatic languages (Burrow 1947a).

(7) **vāśi** 'knife', axe, etc.' RV, Oss. **uaes** 'axe' (← Ir. ***vasa**, M53 s.v.): ← PD ***vāc** 'cut, chip, peel' DEDR 5340 (Burrow 1947a); cf. also PSD1 ****vāy-cci*** 'adze' DEDR 5339, PD ***va(ṇ)ṭi** 'sharp(ness), to sharpen' DEDR 5349, PSD ****vāḷ*** 'sword' DEDR 5376; PD ***mās/mas-** 'sharpen' DEDR 4628, PD ****maẓ-V, *maṯ-Vcc*** 'axe' DEDR 4749 (see Zvelebil 1970: 126 for **m-** ← ***v-**). Note Proto-Elamo-Dravidian ***vac-i** 'plane, remove, cut off' (McAlpin 1981: 105, #77). M53 notes no non-Indo-Iranian cognates.

Notes

1 One of the results of the lack of success in cracking the Indus script is that the subject remains fair game for a wide range of theories: for example, a recent publication (Shendge 1997) identifies the Harappan language as Akkadian, the "earliest known Semitic language."

2 For example, Ahirani/Khandeshi **kanḍolni** 'female members of a wedding party', **kāmoḍyā** 'cotton plants' (Chitnis 1964: index, pp. 190–257); Katkari **cāmblābhuṇḍ** 'a bird', **aḍūsā** 'a plant, *Justicia ganderusa adhenatoda*', **aroḷ** 'a kind of fish', **kas** 'neck', **gojli** 'earthen pot' (Kulkarni 1969); **āhiḍā** 'musk deer', **cāmbāṭā** 'species of bat', **śisuḷ** 'muskrat', **tirgā** 'woodpecker'(Bhili id.), **phesā** (Bhili **phesrā**) 'red-vented bulbul' (Kulkarni n.d.). On the morphological side, Katkari has a distinction between "active" and "passive" forms in the past indefinite and pluperfect tenses, a distinction unlike anything found in Marathi or other major IA languages (Kulkarni 1969: 368–71, 373–5; cf. also Southworth 1976b).

Possehl and Kennedy (1979) hypothesized an ancient economic exchange relationship, which included interbreeding, between the Harappan people of Lothal and local hunter-gatherers during the early second millennium BCE (see also Possehl 1976). It is possible that the "Proto-Bhili" substratum is connected with the hunter-gatherers of that period.

3 Masica (1979: 134) finds that about 30 percent of Hindi agricultural vocabulary, after exhaustive etymological searching, is of unknown origin. Of the 75 words in this category which are found in Turner's CDIAL just 6, or 8 percent, are represented only in Hindi (or in Hindi and one immediate neighbor, Panjabi or Awadhi): 27. **arhaṛ** 'pigeon pea', 31. **guār** 'cluster bean', 78. **ṭiṇḍā** 'squash melon., 93. **caulāī** 'amaranth', 126. **binaulā** 'cottonseed' (as fodder), and **tendū** (see Masica's list, 1979: 76–102). Ten or more items in the list of 75 (13 percent) have skewed geographical distribution, some represented mainly in Eastern languages along with Hindi, and a few mainly in the west and southwest, but the number of cases is not sufficient to enable any conclusions to be drawn.

4 "We can start with the convenient list of Kuiper (1991) … This list has been criticized by Oberlies (1994) who retains "only" 344–358 words, and minus those that are personal names, 211–250 'foreign' words. (fn. 2) … even Oberlies' lowest number would be significant enough, in a hieratic text composed in the traditional poetic speech of the Indo-Iranian tradition, to stand out, if not to surprise." (Witzel 1999b: 3)

5 Witzel (personal communication, 29 January 2003) makes the following suggestions regarding the occurrence of Dravidian words in OIA:

(1) Dravidian words begin in the middle and late RV (as noted above) and increase in number in the post-RV texts;
(2) the late RV-early post-RV period saw a geographical shift to Kuruksetra and the emergence of a new Bharata/Kuru dynasty (Witzel 1995c);
(3) new, non-Aryan elements were incorporated into the Aryan society at this time, probably including people from Dravidian-speaking and Para-Munda-speaking traditions – as evidenced by the occurrence in texts of numerous personal and tribal names – and a thorough reordering of the society took place (op. cit. 9–10);
(4) at this time the Sudra make their first appearance (RV 10.90); possibly these were Dravidian speakers from Bolan/Pirak, Kachi (cf. the mention of Sudroi in this area by Alexander's historians) and possibly Sindh; this name was applied to all of the non-elite, non-Indo-Aryan-speaking groups absorbed by the new society;
(5) if this kind of integration took place before and during the move to Kuruksetra, it would explain why there are many Dravidian words in the texts of this period, but no Dravidian place names in Kuruksetra or to the east of it.

6 An example is Przyluski's derivation of OIA **udumbara** '*Ficus glomerata*, glomerate fig tree', which is based on the assumption that this word contains the element OIA **tumba** 'a type of gourd', a further assumed connection with Marathi **tamburā** 'tanpura, a stringed instrument', and a further implication that these words refer to an Indian caste which is associated with this particular instrument (Przyluski 1926, in Bagchi 1975:157–9). In spite of its complexity, a number of scholars accepted this etymology (see e.g. M53 s.v.), overlooking a straightforward derivation from Dravidian ***uttu-mara** 'fig/date tree' (see **3.22A(3)**).

7 In a number of cases Mayrhofer's later dictionary (M86) is more conservative than the earlier one (M53) in the matter of accepting Dravidian origin – see for example **bilva** in **3.22A(2), atavi, anala**, and **bilva** in **3.22A(4)** – though in at least one case the reverse is true: see **elā** in **3.23(5)**. The point here is not to belittle Mayrhofer's great achievements, but to illustrate the difficulty in weighing conflicting evidence in individual cases. Even if we accept only those items on which Mayrhofer agrees with himself, they would include a number of Rigvedic words, though perhaps none from the earliest books (see the discussion of Witzel in Note 8).

8 For the early RV Witzel accepts only four Dravidian items as even remotely possible (**ukha** 'hip', **phalgu** 'minute, weak', **āṇi** 'lynchpin', and – "very tentatively" – **bala** 'force') (1999b: 30).

9 As noted in **8.2** in the discussion of reconstructing Proto-Dravidian, the non-literary languages Brahui, Malto, and Kudux provide crucial data for the earlier levels of Proto-Dravidian. Since these languages have been for centuries surrounded by languages of other families (Indo-Aryan and/or Iranian) belonging to more advanced communities, there has been much replacement of original Dravidian words in these languages by words from neighboring languages. For this reason, it often seems reasonable to equate words found in later levels of Dravidian with those occurring earlier in OIA sources.

10 The word **kuṭa** also occurs once in RV (1.46.4) with uncertain meaning, possibly 'earthen pot'.

11 If PD had initial ***k-**, this would appear as ***c-** in SDr and as ***k** in CD, and thus would not explain CD words like Pa **cīnd**, Ga **sīndi**; if on the other hand PD had initial ***c-**, it would not be expected to appear as Kudux **k-**. Krishnamurti (2003 §**1.2**) reconstructs this form with an initial ***c-**.

12 Words which appear to have followed this route (Sanskrit → literary Dravidian → non-literary Dravidian) are also excluded from this list: for example, DEDR App. 4 Ka **aggaḷa** 'greatness, eminence, excess' ← OIA **agra** 'top, summit'; in this case the non-literary languages have meanings like 'many, more, extra', probably derived from Te **aggalincu** 'to increase, grow, exert'.

13 Though this entry shows cognates in CD and ND languages, they are likely to be borrowings from Bangla or Oriya **kumṛā** 'gourd' (CDIAL No. 3374).

14 The CD cognates mentioned in this entry are probably borrowings from Marathi **bā̃ḍi** or Oriya **baṇḍi**.

15 Krishnamurti reconstructs retroflex **ṭ ṇ ḷ ẓ** in Proto-Dravidian, contrasting with dental **t n** and alveolar **ṯ l r** (2003 §4.1).

16 It appears to be generally agreed that the first step in this development was the (allophonic) change of the earlier palatal **š** (which had developed out of PIE **s** before **r, u, k, i**) into retroflex, thus avoiding a merger with the new Indo-Iranian **ś** ← PIE palatal ***k** (Burrow 1973a: 18ff.; Deshpande 1978; Hock 1979). This **š** (with an alternant **ź** in voiced environments) developed into retroflex **ṣ** in OIA, though not in Iranian. Subsequently, apical consonants in contact with this **ṣ** developed retroflex allophones, which were at first non-distinctive but later became distinctive as a result of other changes: for example, OIA **nīḍa** 'nest' is derived from a PIE ***ni-sd-os** (compare Gothic

nist, E. **nest**) through the following series of changes: ***-isd-** → ***-išd-** → ***-iźd-** → ***-iẓd-** → ***-īḍ-**, the last change involving loss of the **-ẓ-** with "compensatory lengthening" of the vowel. (See Burrow 1973a for details.)

Hock's view is that this change took place by "polarization" (1979: 47): "if IIr. **š** was to remain distinct from this palatal **ś**, this could be achieved quite naturally by articulating **š** not as a palatal, but as a retroflex" (1975:101). However, certain other outcomes would be equally natural – such as the total merger of old **š** and new **ś**, or the development of **ś** into an affricate **ᵗs** or the like, as in Kafiri (Burrow 1973a:74). But, in fact, none of these things happened in OIA; instead, the old **š** became retroflex, introducing a new phonological component in this limited environment. Thus it is likely that in the variety of proto-Indo-Iranian which was to become OIA, the old **š** had retroflex variants (in some environments) *before* the change of PIE palatal ***k** to **ś**. And such a development would be quite natural if the speakers of this language were in contact with speakers of other languages with phonological retroflexion. For example, a sound like the Dravidian ***ẓ** might have been substituted for the OIA **š** by non-native speakers of OIA. (The Dravidian **ẓ**, which survives in modern Tamil and Malayalam as a distinct phonological entity, and must be reconstructed for PD, seems to have been a "kind of retroflex fricative" (Zvelebil 1970: 148; see also Krishnamurti 1958). It was presumably voiced (though non-distinctively) in most or all positions, and thus would have been the closest equivalent in early Dravidian to the voiced variant **ž** of Proto-IIr (and pre-Vedic) **š** which occurred in ***nižda** 'nest'.) However, the Dravidian ***ẓ** is not the only possibility here; see for example Hock (1993: 97ff).

17 According to this proposal, **r** + dental stop of OIA ended up predominantly as retroflex **ṭ/ḍ,** etc. in the Ashokan inscriptions of the northwest and east, whereas the southwestern inscriptions show primarily dentals; the central area shows both dentals and retroflexes, probably indicating that it is a primary dental area with influence from the east. Hock refers to Turner for evidence that the same developments appear in NIA. The boundaries of these areas coincide closely with a boundary within Dravidian, which marks a division between areas where alveolar **t̲t̲** → dental **t(t)** (west of the boundary), and those in which alveolar **t̲t̲** → retroflex **ṭṭ**. (This is a simplified version of the situation described by Hock.)

If we then assume, as Hock does, that OIA **rt** etc. passed through an alveolar stage on the way to splitting dialectally into **tt** on the one hand, and **ṭṭ** on the other, then the stage is set for the assumption of a widespread common areal development involving Indo-Aryan, Dravidian, and possibly other languages. Though there is no direct evidence for the alveolar development in Indo-Aryan, and though there are other problems (such as the chronology of these changes, which Hock addresses), this proposal does provide a good all-round explanation of the prehistoric linguistic situation – along with suggesting solutions for a number of other questions of historical phonology in Indo-Aryan, Iranian, and Dravidian.

18 Classical Tamil **vit̲t̲i** means 'having caused to hear', that is, 'having said' (causative of the verb **vin̲** 'hear' ← PD ***vin̲** (DEDR 5516). This same meaning is found in the Dravidian particles such as Tamil **en̲r̲u**, and in the Indo-Aryan forms which imitate the Dravidian forms, such as Marathi **mhaṇun**, from **mhaṇ** 'say', and Bangla **bole** (Klaimann 1977). Lest it be thought that such a borrowing is unlikely, it may be noted that the Marathi of Kasargode has borrowed the quotative particle **-ṇṇi/-ṇṇu** from the local Malayalam (see **4.26**).

19 As Witzel (1999b: 28) has pointed out, if words like OIA **tan(ū)**:Avestan **tanū** 'self' and OIA **iti**:Avestan **uiti** (quotative particle) were borrowed from another language, this would imply a close relationship between Proto-IIr. and the source language. The borrowing of grammatical words belongs to what has been called *third-level convergence* in Chapter 4 (see esp. **4.6**), and usually implies a high degree of symbiosis

between the borrowers and the source, with speech-community-wide bilingualism for at least one of the groups. Since the overall number of possible loanwords listed here does not support such a notion, there appears to be a contradiction. This is of course a good reason for those who disbelieve in Dravidian influence at this early period to reject these items as borrowings. However, this apparent contradiction can perhaps be resolved by recognizing the effects of a *convergence continuum* (see **4.3**). Long-term language contact leads to a situation in which some features, and some dialects or varieties of a language, show higher degrees of convergence than others. That there were other varieties of Indo-Aryan contemporary with Vedic is not in dispute: see **2.82**. Some of these varieties were probably more heavily influenced by indigenous languages than the ritual language, and Dravidian influence could be assumed especially for those varieties which were spoken in Sindh, where Witzel assumes Dravidian languages were spoken, at least from the middle Rigvedic period onward (1999b: 32ff., and see the preceding paragraph). The suggestion of earlier contact at the Proto-Indo-Iranian period would put this linguistic relationship back as much as a millennium or more.

20 Many details remain to be worked out in order to make this hypothesis acceptable, particularly on the side of Proto-Dravidian phonology. Though Krishnamurti (2003) reconstructs Proto-Dravidian with a full set of retroflex consonants, some of these may be the result of earlier consonant clusters (see Tikkanen 1988); Hock (1993: 98) points to alternations involving retroflex, dental, and alveolar consonants in Dravidian which have not been taken into account in existing accounts of Dravidian phonology. A second problem concerns chronology: the relevant changes in Indo-Aryan appear to have been completed by the time of the Ashokan inscriptions in the mid-third century BCE, whereas the allegedly related Dravidian changes were still ongoing as late as the ninth or tenth century CE (Hock 1993: 100). As Hock points out, this lag of one thousand years or more may be deceptive, in that the written records in Dravidian languages which show distinct alveolar geminates may be using an archaic spelling; furthermore, the few remaining cases of alternation between retroflex and dental consonants in the Ashokan inscriptions may be the result of lingering alveolar consonants in some areas. In any event, this would not be the only case of a linguistic change which persisted over many centuries in South Asia: note the discussion of the history of the **-l-** past in Outer Indo-Aryan (see **6.2**); cf. also Krishnamurti's discussion of the ancient and still ongoing process of 'apical displacement' in Dravidian (Krishnamurti 1978).

21 It is significant that Colin Masica's mapping of the features of the South Asian linguistic area (Masica 1976: 181) shows that most features do not stop at the borders of South Asia, but extend into many other areas of Asia, particularly Central Asia.

4

THE SOCIAL CONTEXT OF LINGUISTIC CONVERGENCE*

Contents

4.1. Introduction

As noted in **1.32C**, evidence of borrowing of words between two languages generates an inference of contact between speech communities. While the *content* of the borrowed vocabulary often provides direct evidence for the cultural areas of contact, the *quantity* and *types* of borrowed material may provide evidence of the *intensity*, *range*, and *dynamics* of contact. In the present context, *intensity* of contact refers to the degree of social integration among groups in contact (**4.2**); *range or extent* of contact refers to the proportion of each group which is involved in contact, or the numbers and kinds of sociolinguistic situations affected by

contact (**4.3**); the *dynamics* of contact refer to the processes of change in the context of the interrelations among the speech communities in contact (**4.4**).

4.11. Definitions

The term *convergence*, common among linguists working in the South Asian field, is often used interchangeably with *diffusion*, which is preferred for example among the (mostly anthropologically oriented) linguists who have worked on American Indian languages. Since the two terms can express a useful distinction, it would be best to keep them separate. I suggest the following preliminary definitions:

(1) *Convergence* is the gradual reduction of difference between two languages, either in the structure of rules or in the surface structure. (See examples below.) Convergence can be measured in as fine or as gross a degree as desired: a slight statistical favoring of a phonetic variant of language A, when a similar variant is present in language B, or a slight increase in the frequency of a particular order of words or constituents in A, when that order is a characteristic of B, could be said to constitute convergence between A and B as defined here.

(2) *Diffusion* refers to the transfer of a specific feature or features from one language to another. By this definition, any case of diffusion would be a case of convergence, but the reverse does not necessarily apply: for example, two languages can converge without any diffusion between them, as a result of diffusion from a third language. Thus, a number of languages of South Asia share vocabulary items like H–U **isṭeśan** ‘station’, **pulīs** ‘police’, **kār** ‘car’, etc., and the elite varieties of some of these languages have introduced new phonological elements such as the **æ** of Marathi **mæṭ** ‘mat’ and **bæṭ** ‘bat’, etc., as a result of diffusion from English. The same is of course true of many world languages.

For the present purpose we may distinguish two meanings of convergence: *convergence*$_1$ will refer to the general process of reduction of differences, while *convergence*$_2$ will be used to refer to convergence through direct diffusion, which in fact is the sense in which the term is generally used in linguistic area studies, and the sense in which it will be used throughout this work (without the subscript numeral).

Two further distinctions are needed. First, we distinguish between convergence and *shared linguistic change* – since we shall want to speak of convergence among related as well as unrelated languages. A shared innovation (such as the pronominal changes which distinguish the South Dravidian languages from the remaining members of the family, see Krishnamurti 1975) does not constitute convergence as defined here. Nor does a change which spreads gradually across a dialect continuum, such as the change **v** → **b** which has affected the whole area of central and eastern Indo-Aryan (see Masica 1991: 459). It is assumed that the

presence of *bilingual speakers*, who transmit features from one form of speech to another, is a necessary feature of convergence, which distinguishes it from such shared change.

The term *convergence*, as used with respect to South Asia, usually excludes lexical borrowing; that is, it refers primarily to *structural* (phonological, grammatical, or semantic) rather than *lexical* diffusion. In the present context, it will be useful to apply the term in the more general meaning, that is, to include both structural and lexical convergence, for the following reasons:

(1) As will be shown presently, there is a continuum of lexical items ranging from free words (nouns, interjections) through grammatical words (prepositions, conjunctions) to inflectional affixes, such that it is difficult to decide non-arbitrarily where "lexical" borrowing stops and "grammatical" borrowing begins. Certain important measures (see below) can be applied to all parts of this continuum.
(2) While in some cases it is possible to distinguish between the borrowing of an item and the borrowing of a rule or a grammatical process, for the large part of this continuum both are inextricably bound together.

The final definition of convergence, as the term will be used in the remainder of this book, is as follows:

Convergence: the gradual reduction of distance between languages as a result of the diffusion of features through the medium of bilingual speakers. Where desired, *lexical convergence, grammatical convergence, phonological convergence*, and *semantic convergence* can be distinguished as specific subtypes of the general phenomenon.

The transfer of a feature from a speaker's native language into a non-native language will be called *carry-over*; (2) the transfer from another language into a speaker's native language will be termed *adoption*. This distinction of course can only be applied where it is possible to distinguish between native and non-native language (see **4.32**).

4.2. Intensity of language contact

Sherzer and Bauman (1972) demonstrated that two different types of linguistic diffusion can be used to make different types of inferences about prior cultural contacts. They suggested that, on the one hand, the sharing of "areal patterns in semantics, grammar, and phonetics indicate intimate face-to-face contact between whole communities," whereas areal patterns in lexicon (without the sharing of structural patterns) indicate less intimate levels of intergroup contact – and in some cases may even reflect independent responses to

shared ecological factors, thus not necessarily implying any direct contact at all (1972: 144).[1]

While accepting this as a general characterization of the kinds of contact situations which produce these different types of linguistic diffusion, we may ask what further steps might be taken in two directions of interest:

(1) generalizing Baumann and Sherzer's conclusions beyond native North America and the types of social organization found there;
(2) examining the possibility of finding more detailed relationships between the types of linguistic features diffused and the social characteristics of the contact situations involved.

The initial data for this investigation come from a dozen South Asian languages, which have been studied with varying degrees of intensity. The conclusions drawn from these languages will be tested against a number of cases drawn from other areas of the world.

4.21. Correlations in the linguistic data

Before beginning to look for correlations between kinds of linguistic convergence and kinds of contact situations, it will be necessary to find ways of quantifying or ranking the linguistic manifestations of convergence, so that we have some reasonably consistent way of identifying *degrees* of convergence. Several criteria suggest themselves for this purpose – including, on the side of lexical borrowing:

(1) the *grammatical categories* of borrowed words – for example, it seems to be generally true that concrete nouns are borrowed with relative ease, whereas function words tend to be borrowed only in later stages and/or in more "intimate" types of contact – a hypothesis which needs testing;
(2) *percentages* of lexical borrowings – higher percentages presumably reflecting longer, more intimate, or more widespread contact;
(3) the distinction between *cultural vocabulary* and *core vocabulary* – the diffusion of the latter presumably implying greater bilingualism and more intimate social contact than the former.

When it comes to the convergence of phonological and grammatical patterns, quantification poses some problems which will be discussed later. For the moment, we may assume that some sort of *relative degrees of structural convergence* can be defined, which can then be used to seek correlations with measures of lexical convergence on the one hand, and with types of contact situations on the other.

4.22. Lexical convergence

Convergence whose salient manifestation is the *transfer of overt morphemes* will be termed lexical convergence, whereas convergence which manifests itself primarily as *rule change* will be called structural (grammatical or phonological) convergence (see Weinreich 1974: 29–30 for a related distinction). In lexical convergence – whether it involves phrases, content words, function words, or inflectional affixes – the forms which diffuse generally belong to comparable categories in both languages (Weinreich 1974: 32), or fit into pre-existing slots or categories in the borrowing language. Convergence which involves the creation of new grammatical categories is necessarily (by definition) grammatical convergence.[2]

4.23. The morpho-lexical continuum

Various writers have suggested the notion of a hierarchy of transferability, or "scale of adoptability," which would make it possible to rank lexical borrowings both by grammatical category and in terms of the ease with which they diffuse across language boundaries (e.g. Haugen 1950: 97). I do not know of any attempt to test this notion. Table 4.1 presents a number of cases from South Asia, in which the borrowings are classified according to the traditional parts of speech. (For the purposes of this table, lexical items have been classified in terms of their role in the *borrowing* language, regardless of what role they may play in the source language. Thus for example, when English verbs are borrowed into South Asian languages, they frequently appear as *verbal complements* (second column in the table) rather than as verb stems in the borrowing language; that is, they are used in construction with an "all-purpose" verb like Hindi–Urdu **karnā** 'to do' or **honā** 'to become', or the like: for example Hindi–Urdu **fel honā** 'to fail', **fon karnā** 'to telephone', etc.)

Table 4.2 shows the same data reorganized, along with comparable data from several other areas of the world. The first part of this table is derived from Table 4.1 first by lumping the smaller categories into four main ones: (1) substantives, (2) other content words – mainly verbals and modifiers, (3) function words, (4) inflectional affixes; and second by dropping interjections, phrases, and derivative affixes, either because they are not diagnostic for the present purpose, and/or because they are not adequately reported. Interjections and phrases which can be used as independent utterances appear fairly frequently in lists of borrowings, and experience in South Asia indicates for example that some English phrases (such as **Good morning, sir/madam** or **Thank you**) are found even in the speech of uneducated monolingual speakers of local languages, especially in urban areas. There are, on the other hand, types of English phrases which seem to be more symptomatic of the speech of educated bilinguals (such as **You know…**, **By the way…**, **As a matter of fact…**, **If you don't mind…**, **What rubbish!**). Thus a distinction between these two types might be useful, though our sources do not generally provide the necessary data to fit them into our scheme.

Table 4.1 Lexical borrowings by grammatical category in selected South Asian languages

Languages[1]		*Grammatical categories*															
Borrowing language	*Source language*	*Content words*					*Function words*									*Affixes*	
		Substantives		*Other*													
		Noun	*Verb Comp.*	*Adjective*	*Verb*	*Numeral*	*Conjunction*	*Adverb*	*Postposition*	*Num. Classifier*	*Quote marker*	*Aux. verb*	*Pronoun*	*Interjection*	*Phrases*	*Derivational*	*Inflectional*
Saurashtri	*Tamil*	×		×	×	×		×				×				×	
Brahui	Iranian	×	×	×	×	×	×	×					×				×
Konda	Indo-Aryan	×	×														
Konda	*Telugu*	×	×	×		×						×	×				
Malto	Indo-Aryan	×	×		×	×				×							
Hindi–Urdu[2]	Persian	×	×	×	×	×	×	×	×				×	×	×		
Gondi	Indo-Aryan	×	×	×	×	×	×		×					×			
Kasargod Marathi	*Dravidian*	×	×		×	×					×			×			
Tamil	English	×	×											×			
Kudumbi[3]	*Malayalam*	×	×			×								×		×	
Malayalam[4]	English	×	×	×		×	×	×	×	×			×	×	×		

Notes
1 Italicized languages are Dravidian; others are Indo-European.
2 All categories listed here are found in the 1,000 most frequent words listed in Ghatage (1964).
3 Data for this language are fragmentary.
4 "Code-switching" style (see **4.31**).

Table 4.2 Lexical borrowings by grammatical category in selected world langauages

Languages		*Grammatical Categories*			
		Content words			
Borrowing language	*Source language*	*(1) Substantives*	*(2) Other*	*(3) Function words*	*(4) Inflectional affixes*
Saurashtri	Tamil	×	×	×	×
Brahui	Iranian	×	×	×	×
Kudumbi	Malayalam	×	×	?	×
Konda	Telugu	×	×	×	
Malto	Indo-Aryan	×	×	×	
Hindi–Urdu	Persian	×	×	×	
Gondi	Indo-Aryan	×	×	×	
Kasargod Marathi	Dravidian	×	×	×	
Konda	Indo-Aryan	×			
Tamil	English	×			
Meglanite Rumanian	Bulgarian/ Macedonian[1]	×	×	×	×
Albanian	Greek	×	×	×	
Bulgarian	Rumanian	×	×	×	
Rumanian	Slavic	×	×	×	
Greek	Turkish	×	×	×	
Greek	Albanian	×	×		
Rumanian	Greek	×	×		
Turkisk	Greek	×			
Rumanian	Albanian	×			
Ngandi[2]	Ritharngu	×	×	?	×
English	French	×	×	×	
Ateso[3]	Swahili	×	×	×	

Notes
1 Weinreich 1974: 32; Sandfeld 1930.
2 Heath 1981 (items are shared; source is unspecified).
3 Scotton and Okeju 1973.

A derivational affix can be said to have been borrowed when enough words containing it have diffused into a language that the affix becomes productive and is used with native stems. Thus, Hindi–Urdu has many Persian- or Arabic-derived words containing a negative/privative prefix **be-**, such as **bevaqūf** 'brainless, stupid', **beizzat** 'dishonor, disrespect', **beīmān** 'dishonest', **beśarm** 'shameless' which can also be combined with inherited Indo-Aryan words, such as **bekām** 'useless' (= **be + kām** 'work' [←OIA **karma**]), **bekār** 'unemployed' (= **be + kār** [←OIA **kārya**]). The point at which this type of hybrid formation begins to occur is no doubt an important milestone on the historical convergence continuum, but again our sources do not generally provide adequate data on this point.

The data presented in Table 4.2 would appear to support, at least tentatively, the view that the four categories used here can stand as an *implicational* continuum

reflecting the progress of linguistic convergence; that is, a mark in any column implies that the column to the left is also filled, but not the reverse. There are, however, four major reservations to this statement, which is why it must be regarded as tentative:

(1) The categories may not be suitable for a truly cross-cultural approach – for example, for languages in which categories such as substantive, or derivational affix, are not meaningful.[3]
(2) There is the matter of resistance to the diffusion of some categories of words: for example, Heath 1978 mentions that one category of verbs is not easily diffused among the languages he studied in northern Australia. As noted earlier, English verbs are not often borrowed *as verbs* in South Asian languages. Thus the scheme presented here may need to be modified in particular cases to take such structural factors into account.
(3) There are two cases in Table 4.2 (Kudumbi and Ngandi–Ritharngu) which show borrowing of inflectional affixes though they have not borrowed function words, according to the available descriptions. In the first case it is possible that adequate data would fill this gap; and perhaps in the second case also, since Heath's focus was on the quantum of borrowing in certain semantically defined categories, and not on the *grammatical* categories of the borrowed items.
(4) The fourth reservation must be, of course, that more data are needed from other parts of the world before any such scheme can qualify as a sociolinguistic universal. On the other hand, there is nothing very new about this proposal, and perhaps nobody would be very surprised if it turned out to be (more or less) true.

This proposed sequence would seem to reflect a progression from minimal, or partial, bilingualism to a more widespread and higher-level bilingualism. Thus by all accounts, concrete nouns (apart from some minimal greetings and polite phrases) are the first and most frequent lexical items learned in a new language, the names of concrete objects being relatively easy to use with a minimal knowledge of the source language. Furthermore, the substantive category is typically open-ended by contrast with other categories. Borrowing other content words such as adjectives or verbs would imply a greater knowledge of the sentence structure of the source language than borrowing nouns; borrowing function words would imply a still greater knowledge, both on the part of the user and of the hearer.[4] Furthermore, these "higher" categories of borrowings often involve words or phrases which duplicate, at least partially, items already present in the borrowing language – as opposed to borrowed substantives, which most frequently refer to salient cultural items in the source language. Thus the sequence of grammatical categories will be seen to overlap partially with the distinction between cultural and non-cultural vocabulary to be taken up below.

The borrowing of inflectional affixes is a rather rare phenomenon in the general picture of linguistic convergence. Weinreich (1974) mentions only a single case, that of verb endings which have diffused from Bulgarian into Meglenite Rumanian, spoken in Greek Macedonia (Weinreich 1974: 32). In Brahui, which has been studied in detail by Emeneau, there is only one rather uncertain case (1962a: 56–8), in the midst of massive lexical borrowing from Iranian languages which has nearly obscured the Dravidian origin of Brahui (see **1.22A, 2.41**). The only case mentioned in Heath's study of Arnhem Land occurs between two languages which share as much as 75 percent of their vocabulary in some domains, largely as a result of recent convergence. The other cases mentioned for South Asia, those of Kudumbi and Saurashtri, involve languages which also show very high levels of grammatical convergence. Virtually all members of the speech communities in question are bilingual, and have been so for generations. Thus, to anticipate the discussion of social factors in linguistic convergence below, we can say that when inflectional affixes appear in diffused material they suggest both the presence of large numbers of borrowings belonging to the first three categories of Table 4.2, and also pervasive high-level bilingualism or diglossia for at least one of the groups involved.

4.24. Quantum of lexical diffusion

The available studies of lexical convergence generally look at the impact of borrowing on a single language, ignoring the question of reciprocal influence. Heath's study of the languages of Arnhem Land (1978) is an exception, but there the source languages have not been identified – a difficult matter, since all these languages appear to be ultimately related to each other. Nevertheless, the high figures Heath gives for shared vocabulary in Ngandi and Ritharngu, varying from around 25 percent to over 75 percent (depending on the semantic category of the lexicon) are unusual, and even in core vocabulary he gives a figure of 28 percent (see Table 4.3) which seems to be largely the result of recent borrowing.

By contrast, the other cases shown in Table 4.3 have uniformly low figures for core borrowing, but show considerable variation in non-core vocabulary. The figures for Kasargod Marathi are probably fairly typical for South Asian languages spoken by small isolated groups. Hindi–Urdu, which is not included in the table, would show from 4 to 7 percent of borrowings from Persian in its basic vocabulary, depending on whether one is dealing with the Urdu or Hindi end of the continuum. In the most frequent 1,000 words in a published Hindi frequency list (Ghatage 1964), 8 percent are of Persian/Arabic derivation.[5] The figures for Bella Coola, taken from Newman's (1956) study of Bella Coola lexical borrowing, show a large difference in the amounts derived from different sources: 57 percent of "ecological" vocabulary from Wakashan languages, and only 3 percent from Athapaskan languages, though there was close contact with languages of both groups (see further discussion in **4.28**).

Table 4.3 Comparison of grammatical and lexical diffusion for selected world languages

Language	*Grammatical features*	*Lexical level*	*Percentage of borrowing*	
			Core	*Non-core*
Saurashtri	6/7	4	—	—
Dakhini	4/7	4	—	—
Kasargod Marathi	4/7	3	.03	.18
Kupwar Marathi	3/7	3	—	—
Albanian	4/4	3	—	—
Bulgarian	4/4	3	—	—
Macedonian	3/4	3	—	—
Serbo-Croatian	2/4	2	—	—
Rumanian	1/4	2	—	—
Ngandi-Ritharngu	?	4	.28	.50+
Bella Coola (←Wakashan)	14	?	.03	.57
Bella Coola (←Athapaskan)	6	?	—	.03
English (←French)	?	3	.03	.5?

Notes

1 For South Asian languages the number in the second column indicates the number of Dravidian features (out of a possible 7) which have been adopted by the language to the 2nd or 3rd degree; for the Balkan languages the number of Greek features present (out of a possible 4); for Bella Coola the number of Northwest area features shared with the language family in question.

2 Lexical level refers to the highest category in which borrowings occur (see Table 4.2). For the Balkan languages, the figures refer to lexical items of Greek origin.

4.25. Phonological convergence

Though this is an important aspect of the subject, it is only possible to mention the following (perhaps rather obvious) points here:

(1) There are often conspicuous differences among social segments of a speech community with regard to the diffusion of phonological traits. For example, in spite of centuries of French influence on English, only a minority of (highly educated) English speakers have the phoneme **/ź/** in initial position (as in /źánrə/ **genre**). Similarly, the Marathi vowel **æ** (as in **mæṭ** ‘mat’) is largely restricted to those Marathi speakers who have learned English through formal education, and even some of these speakers may occasionally say /meṭ/ [me:ṭ] or /myaṭ/ [myæṭ] (see Kelkar 1957).

(2) The *loss of native distinctions* (such as the loss of the distinction between aspirated and unaspirated stops in Kasargod Marathi, or the loss of some retroflex consonants in Brahui, the loss of the dental–retroflex distinction in the Gypsy languages) appears to be an indicator of a high level of convergence, that is, usually implies the presence of lexical borrowing including levels 1 and 2, and perhaps 3.

4.26. Grammatical convergence

In spite of recent advances in grammatical theory and linguistic typology, it is difficult to find any rational method for ranking grammatical structures which would be in any way comparable to the ranking of lexical items by grammatical categories. The most useful way of approaching this matter is perhaps that employed by Sherzer in his study of linguistic and cultural areas in native North America (1976). In this study, Sherzer selected an inventory of features which occur frequently in the literature on diffusion, and used this inventory as a basis for estimating the extent of shared features in each area. For South Asia a similar method may be used. In looking at the convergence between Indo-Aryan and Dravidian languages, we find that these two groups of languages are widely divergent in structure in many respects. For example:

(1) Indo-Aryan languages have grammatical gender, whereas Dravidian languages have "natural" gender (i.e. inanimate objects are mostly neuter, male creatures masculine, and female creatures feminine);
(2) Indo-Aryan languages have various forms of ergative construction, which are completely absent in Dravidian;
(3) the structure of number names in the two families is different (e.g. 35 = "five-thirty" in Indo-Aryan, but "three-ten-five" in Dravidian – see Pandit 1974);
(4) Dravidian has historically only left-branching relative clauses, whereas Indo-Aryan has both left- and right-branching; etc.

In cases of contact between languages belonging to these two families, some or all of these features undergo change to a greater or lesser extent. In order to quantify these changes, I propose the following system:

(1) *First-degree convergence* involves change in a grammatical pattern which affects only the surface structure manifestations (e.g. relative frequency of a variant, or stylistic constraints on the occurrence of a form) without any major modification of the underlying rule structure. Thus an increase in the frequency of copula deletion, or a variation in noun gender, would qualify as first-degree convergence (provided, of course, that they can be shown to be responsive to language contact).
(2) *Second-degree convergence* involves change beyond the first degree, thus necessarily involving some modifications of underlying rules, but stopping short of obliterating the distinction between two languages with respect to the rule subsystem involved. Thus widespread fluctuations in gender of nouns, or systematic loss of the copula in some but not all environments (as in the case of Kasargod Marathi), would qualify as second-degree convergence.
(3) *Third-degree convergence* involves total or nearly total convergence of languages with respect to some feature or subsystem. Thus for example,

> Saurashtri has lost grammatical gender and now has the same type of "natural" gender found in Tamil; the rules for copula deletion in Saurashtri also appear to be the same as in Tamil (Pandit 1972: 14–15).

In Table 4.3, the ranking of South Asian languages in terms of grammatical convergence (column 1) has been accomplished by giving the number of features (out of a possible 7) which show second- or third-degree convergence. For the Balkan languages, the figures in both columns represent the influence of Greek on the five languages listed there: out of a possible four features which one author (Sandfeld) considers to be of probable Greek origin, these languages show a range of 1 to 4. For Bella Coola, I have shown the features shared with Wakashan (14) and with Athapaskan (6).

4.27. Linguistic correlations

We may note a general correlation between columns 1 and 2 of Table 4.3, that is, between the level of grammatical convergence and the categories of lexical items diffused. The most conspicuous exception to this statement is the case of English and French, for which it is difficult to demonstrate grammatical convergence of the type discussed here.[6] The case of Ngandi and Ritharngu in Arnhem Land, Australia (Heath 1978) shows a number of cases of probable grammatical convergence, though perhaps not in proportion to the indices of lexical convergence, whereas this relationship is almost reversed in Kasargod Marathi and other South Asian cases, which have high or moderate levels of grammatical convergence with fairly low levels of lexical convergence. Thus we can say tentatively that lexical and grammatical convergence are (at least partly) independent processes, that is, they represent separate and independent continua. It seems probable that phonological convergence will turn out to be linked to grammatical convergence and independent of lexical convergence: note, for example, the minimal effects of French on English phonology, and of Turkish on the phonology of the Balkan languages.

4.28. Linguistic diffusion and the social conditions of language contact

On the basis of these (rather tentative) conclusions, can we say anything about the *social* factors which distinguish high levels of convergence from medium levels, or medium from low levels? One point that is immediately apparent is that while medium-to-high levels of lexical convergence (up to lexical level 3 of Table 4.2, and up to 50 percent or more of *non*-core borrowings) can result from long periods of military and/or cultural domination, a more intimate type of contact is apparently required for even low levels of grammatical convergence, or for lexical convergence on level 4 (affix borrowing), or for higher levels of core borrowings. Thus, the impact of 500 years of Turkish rule in the Balkans reached lexical level 3, but produced hardly any core borrowings, and little if any grammatical convergence. Similar results were produced by the impact of Islam and Arabic on

Iranian, the impact of Iranian and Islam on Hindi–Urdu, that of Sanskrit and classical Hinduism on Southeast Asian languages, and that of Franco-Roman culture on English (see **4.27**). Two hundred years of British rule in India may have drastically altered the region's economic and political structure, but in lexical convergence hardly produced more than level 1 – though in some elite styles of language this may rise as high as level 3, as shown later.

By contrast, cases of medium or high grammatical convergence all involve the *symbiosis* of two ethnolinguistic groups, that is, an economic-cum-cultural dependence which is part of the daily routine of life for at least *one* of the groups. In such situations, all or almost all members of at least one group generally show some degree of bilingualism or diglossia. Looking at the extreme cases, those showing lexical level 4 (affix borrowing) and a high degree of grammatical convergence, something more can be said. Of those studied here, Saurashtri, Brahui, and Kudumbi are isolated languages surrounded by a majority of speakers of very different languages. This is also true of Meglenite Rumanian (spoken in the Macedonian region of northern Greece) and Istro-Rumanian (spoken on the peninsula of Istria, south of Trieste in Yugoslavia), the only cases of Balkan languages for which the borrowing of inflectional affixes is reported. It seems also to be true of Bella Coola, isolated from other Salishan languages and interspersed with villages of the Bella Bella (Wakashan) tribes.

Some of these cases appear to show some weakening of the link between language and ethnic identity. The Kudumbi in South India, for example, were beginning to lose their language in the 1970s, and school-going children at that time had at best a passive command of it.[7] It is reported that speakers of isolated Rumanian dialects have no awareness of the existence of a large Rumanian nation, and until recently the speakers of Saurashtri in South India appear to have been unaware of the existence of a Gujarati-speaking state in North India. The difficulties of correlating linguistic repertoires and ethnic identities among the Brahui have been shown very clearly by Emeneau (1962a: 52). Of the languages studied by Heath in Australia we may note that one is recently extinct (i.e. was on the way to extinction at the time it was recorded), and several others have been reduced to very small numbers of speakers, mostly settled at missionary outposts. Many of these people now use a pidginized variety of English as a *lingua franca*.

The phenomenon of affix borrowing has been commented on by Gumperz and Wilson, who found, in their 1971 study of a village in an Indo-Aryan/Dravidian border area, that when their informants listened to recordings containing instances of this phenomenon, the informants did not reproduce the utterances readily, and seemed to consider them aberrant – even though such utterances did occasionally occur in spontaneous speech. Gumperz and Wilson concluded from this that affixes are seen as a fundamental part of what distinguishes one language from another, and thus to mix up affixes is to blur the distinction between languages. In general, the varieties of language spoken in the situation they studied did not produce lexical convergence above level 3, which may be related to the fact that this is a border situation, rather than the situation of an isolated language.

The figure for grammatical convergence in Kupwar Marathi is also the lowest for the South Asian languages listed in Table 4.3.

In South Asia, groups speaking isolated languages tend to establish themselves as separate units in the local socioeconomic hierarchy, often with their own ecological niche, much as groups of immigrants have historically established themselves in certain occupations in the United States.[8] Such small groups often retain their language for long periods of time, though cases of language loss do occur (cf. the Kudumbi case cited previously, where language loss may have been accelerated by attendance in Malayalam-medium schools). The fact that these small groups often show very high levels of grammatical convergence, along with low to moderate levels of lexical convergence, perhaps relates to the "separate but *un*equal" mode of life under the caste system in South Asia – that is, the coexistence of the need for economic cooperation with the tradition of separateness of ethnic life (particularly of marriage and commensality), which is sanctified by the theory of caste with its rules of pollution (see Mencher 1974, 1976). As noted elsewhere, caste-based lexical differences tend to be concentrated in semantic areas dealing with such intracaste activities as food and ritual behavior, whereas they are much less marked in areas such as agriculture which involve intercaste cooperation (Southworth 1979a).

Intermarriage is a factor often invoked in discussions of bilingualism and linguistic convergence. Heath, for example, notes that the rates of intermarriage among the various groups he studied appear to correlate well with levels of shared vocabulary. Scotton and Okeju, in their study of borrowing in Iteso (1973), also suggest the potential importance of intermarriage. More than one student of the subject has speculated on the effects on a child's speech of having parents with different mother tongues. However, Newman's (1956) study of Bella Coola demonstrates that intermarriage *per se* is not a *sufficient* criterion, since the Bella Coola intermarried extensively with both Athapaskan-speaking and Wakashan-speaking groups, but show extensive lexical and grammatical convergence only with the latter. The South Asian cases, on the other hand, demonstrate that intermarriage is not a *necessary* criterion, since intermarriage is rare among the groups discussed here.

4.29. Inferences about intensity of contact

The preceding discussion has examined possible correlations between types of linguistic diffusion and types of culture contact in a variety of regional and cultural settings, and has led to certain generalizations which deserve to be tested against a still wider cross-cultural sample, in order to assess their validity as "universals of linguistic convergence". Briefly stated, these generalizations are as follows:

(1) Lexical and structural convergence represent two essentially separate continua.

(2) Levels of lexical convergence, measured in terms of grammatical categories 1 to 3 and/or in terms of percentages of items in non-core vocabulary, are responsive to *duration of contact*; thus, 500 years of Turkish rule in the Balkans produced higher levels of lexical borrowing than 200 years of British rule in South Asia.
(3) Structural convergence of any significant level, or lexical convergence of level 4 (i.e. borrowing of inflectional affixes), or high levels of core borrowing, reflect *symbiotic* relationships between ethnolinguistic groups in which the borrowing group identifies itself as part of the same society as the other group. Thus the length of time for Turkish in the Balkans is comparable to the period of Saurashtri contact with Tamil, but the latter case shows much greater grammatical convergence.
(4) Extreme levels of the types of convergence mentioned under point (3) imply at least a partial loss of the separate ethnolinguistic identity of the borrowing group. This is clearly an area requiring further investigation by linguistic anthropologists.
(5) While intermarriage and intrafamilial bilingualism are presumably important social mechanisms promoting linguistic convergence, they are neither sufficient nor necessary explanations from a cross-cultural point of view.
(6) Though cultural identification between groups is a *concomitant* feature of high levels of linguistic convergence, it is probably not an adequate *explanation* of it. *Economic interdependence* would seem to be a more potent, and more general, factor.

4.3. Range or extent of language contact

The descriptions from which the above-mentioned examples were drawn generally do not mention differences *within* languages, implying tacitly that borrowed elements are available to all members of the borrowing speech communities. However, differences among social groups speaking the “same language” are often as great as the differences among languages discussed earlier. If linguists have learned one thing from the past few decades of studying sociolinguistic variation in contemporary societies, it is that there is no language without variation. This is as true of languages in contact as it is of other languages: not only is there great variety in language contact situations, but in any single situation different segments of a speech community potentially respond differently to contact situations. An understanding of such variation in contemporary language contact situations, and of the linguistic and social processes involved, can be of great importance in helping the historical linguist to make accurate interpretations of ancient linguistic data.

Many discussions of borrowing or convergence simply list lexical items, phonological features, or grammatical features which have diffused from one language to another without any indication of the *extent of use* of these features – which may vary in a number of dimensions, including (1) the styles of speech

in which they occur, (2) the types of speakers who use them (in terms of socioeconomic status, bilingual competence, education, etc.), and (3) the regions (both geographical and in terms of the urban–rural continuum) in which they are current. Ideally, if we hope to understand the ways in which features are transferred from bilingual to monolingual speakers, we would like to know, for any contact situation at a given time, which features are used by which speakers in which types of situations. Unfortunately, the large majority of treatments of this subject hardly mention any of these points. Thus, for a given list of features said to be diffused from language X to language Y, we often cannot be sure whether there is any single socio-linguistic context in which they all may be found to occur.

4.31. The convergence continuum

Cases of differences within a speech community in the diffusion of phonological features (such as those mentioned in **4.25**) are familiar to linguists. The same kind of variation occurs in the use of borrowed lexical items: for example, uneducated monolingual Malayalam speakers use English words like **bassu** 'bus' or **rōḍ** 'road' in their Malayalam, but do not produce bilingual utterances like those given in **4.31** B(3). The same kind of variation affects different *styles* of speech: thus in India, formal speech in the indigenous languages generally avoids English or other foreign lexical items wherever possible, even if this requires the use of recently coined, infrequently used words which may have an artificial ring to some hearers. Such differences occur even in illiterate communities: among the Zuñi of New Mexico, for example, English words which are common in ordinary Zuñi speech are avoided in religious contexts (Newman 1956 [1964]).

Grammatical convergence also appears to vary in a similar way according to social factors such as socioeconomic class, education, and degree of exposure to native or fluent speakers of the source language, as well as variables relating to the situation: formal–informal, traditional–modern, etc. Thus Indians who have learned English without benefit of formal education consistently fail to apply the Noun–Aux reversal rules in English questions, producing utterances like **What you are doing?** (instead of **What are you doing?**). Those who are educated in English may apply these rules correctly and consistently in writing and formal speech, but sometimes use the non-standard variants in conversation [author's observation].

In such cases we can observe a range of usage of many borrowed features which we can call the *convergence continuum*. The shape of this continuum, in terms of linguistic features and social situations, will differ from one case to another, though certain general elements can be specified. One is the presence of speakers with different *kinds* of linguistic competence. Given two languages (A and B) in contact, and assuming that it is possible to distinguish between native and non-native speakers of each language (see **4.32**), then we would need

to recognize the following six possible types of speakers (shown in Figure 4.1):

(1) A^0 = monolingual speakers of A who have no direct contact with speakers of B or with A–B or B–A bilinguals;
(2) A^1 = monolingual speakers of A who have contact with A–B or B–A bilinguals;
(3) A^2 = native speakers of A who are bilingual in B;
(4) B^2 = native speakers of B who are bilingual in A;
(5) B^1 = monolingual speakers of B, parallel to A1;
(6) B^0 = monolingual speakers of B, parallel to A_0.

In some situations, of course, one or more of these types may be absent or, if present, may make no contribution to the outcome: for example, where A is the language of a small immigrant group, the categories B^1 and B^2 may be non-existent or negligible.

I am not aware of any documentation of such a "convergence continuum", and therefore I provide here some data on the relationship between English and the south Indian languages Malayalam and Tamil, based on a study conducted in 1978–79 (see Southworth 1980). The first set of examples (A) shows the use of Tamil or Malayalam features in English, the second set the use of English features in Malayalam or Tamil. (E = English, M = Malayalam, T = Tamil; S = Malayalam or Tamil.) In both of these cases, we are presumably dealing with an implicational continuum, in the sense that any feature appearing at one stage is available at the next (higher-numbered) stage, but not necessarily the reverse.

A. Tamil/Malayalam features in South Indian English

(1) Fully integrated E lexical items such as *curry* and *mongoose* are used by all E speakers and by S^2 speakers in all types of interactions.[9]
(2) "Indian English" lexical items such as *lady-fingers* 'okra', *brinjal* 'eggplant, aubergine', *tiffin* 'lunch, snack' are used by E^1, E^2, and S^2 speakers freely in all appropriate interactions and situations.

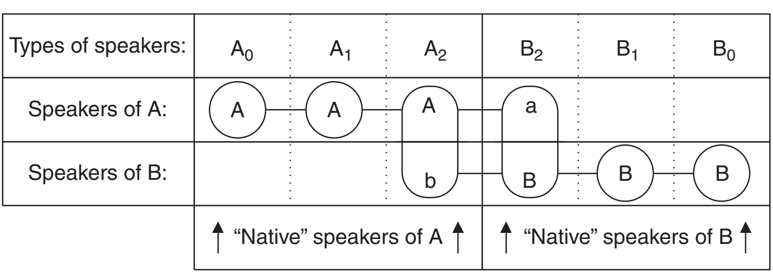


Figure 4.1 Types of speakers in contact situations.

Notes
1 Circles/ellipses represent the linguistic repertoires of groups of speakers.
2 A, B = mother tongue (MT); a, b = other tongue (OT).
3 Lines connecting circles/ellipses indicate contact between groups of speakers.

(3) Retroflex stops for alveolars, S intonations (as well as hand and facial gestures) are used by S^2 speakers when speaking English, with variations depending on the speaker's level of competence (used less by speakers with greater exposure to native E speakers and fluent S^2 speakers of E) and identity of hearers (used less, for those with the requisite competence, with E hearers than with S hearers). E^2 (and even E^1) speakers also use these features either regularly or sporadically (inadvertently or jocularly) depending on the speaker's level of acculturation and the hearer's identity and status.

(4) The following features (and others not listed here) are used primarily by S_2 speakers with either S^2 or E^2 hearers, or with speakers of other Indian languages (with constraints as in the preceding), and perhaps also by highly acculturated E^2 speakers:

- article deletion: 'You are coming on tenth?'
- **yes/no** reversal:
  'Don't you have a copy?' – 'Yes.' (= I don't have a copy.)
  'Aren't you coming?' – 'No.' (= I *am* coming.)
- **V-ing** with stative verbs: 'He's not liking it.' (= He doesn't like it)
- **isn't it** as all-purpose question tag: 'You're coming, isn't it?'
- "enunciative **u**": Turn left. [lɛftɨ] (see Bright 1972, 1975)
- **itself** as all-purpose emphasizer: 'You spoke to the director itself?' (= himself, personally)
- object deletion: 'He doesn't want [tea, food, etc.].'
- word order: 'What he is doing?', 'By train if you go…'
- transitive–intransitive interchange: 'That letter got yesterday' (= …was received…), 'Water is not obtaining here' (= …cannot be obtained…)

(5) The following features were used only among S^2 speakers using "code-switching style," a highly informal interactional style:

- intra-sentential switching between E and S: 'NĀLAYKKI (tomorrow) you are coming, ILLEYĀ?' (isn't it?);
- 'AVAN̲ VANTĀ (if he comes), then I will also come.'
- use of function words and particles from S: 'I am coming-ALLO' (you know)
- 'IPPOM (now), children VANTU (as far as they are concerned), they are learning Tamil'
  'Ō (yes), I can express myself'

B. English features In Tamil and Malayalam

(1) Fully integrated E words in S include substantives, most of which are related to modern technology, customs and institutions introduced during the British colonial period such as **bus**, **taxi**, **motor**, **windscreen**, **road**, **school**, **college**, **doctor**, **x-ray**, **court**, **judge**, (electrical) **current**, **light**, **bank**, **account**, **manager**, **shop**, **shopping**, **cinema**, **pant(s)**, **frock**, **shirt**, **Mummy**, **Daddy**,

brother, **sister**, etc. In addition, certain E verbs appear in S as verb complements with a following verb, as in M **transfer** ĀKU 'be transferred', **register** CEYYU 'register', T **repair** PAṆṆU 'repair'. Such lexical items, as well as phrases like **employment exchange**, **school fees**, **First-class**, **prime minister**, **film star**, are used freely in appropriate situations by S^0, S^1, S^2, and E^2 (E–S bilingual) speakers of all socioeconomic and educational levels.

(2) A much wider range of E nouns and verbs is found in the speech of S^2 speakers who are educated in English, especially those who have lived abroad or have had extensive contact with foreigners or S–E bilinguals. Examples from our data include items such as **habit**, **family**, **language**, **interest**, **craze**, **short term**, **maintenance**, **attitude**, **complaint**, **increment** 'pay increase', **future**, nominal plurals such as **children**, **higher studies**, **marks** (school grades), **relatives**, **family members**, **subjects**, **special programs**, and verbs like **continue** CEYYU, **isolate** CEYYU, **use** CEYYU, **supply** CEYYU, **type** CEYYU 'to typewrite'.

(3) In code-switching style (see before), English-educated S^2 speakers use almost the whole range of English grammatical categories in utterances which are otherwise in Tamil or Malayalam. The only E grammatical categories not found in these contexts were pronouns, auxiliary verbs, and affixal morphemes, but note the use of **-ly** adverbs like **mostly** and **occasionally**. Examples:

ATU **conscious** ĀYIYIRUKKUKAYILLA
'It wouldn't be done consciously'

ATU **sure** Ā
'That is sure'

PINNE **hopeless** ĀYI
'So it was hopeless'

Because ENIKKU HINDIYŌṬU KUṬUTAL EṬUPPAM **right from childhood**
'Because I had a lot of exposure to Hindi right from childhood'

That is, HINTI-NNU COLLAPPAṬṬATU **roman alphabet-**LE EZUTURATU**, not in** HINDI LIPILE ALLA
'That is, what is called Hindi is written in the Roman alphabet, not in Hindi script'

IPPOM VŪṬṬILU REMPA **homely-**YĀ UḶḶA **topics-**ṆṆA **mostly** TAMIḺ TĀN VARUM
'Now at home, [if we are talking of] homely topics, mostly Tamil comes' [i.e. we use mostly Tamil]

Sixty-four-IL
'in [19]64'

Association-NU UḶḶE **for** ŌṆAM **and Christmas celebration-**NNU**, only** ŌṆAM **and Christmas celebrate** PAṆṆUM'

'In the association for Onam and Christmas celebrations, we only celebrate Onam and Christmas'

Occasionally INGLISH UPAYOGIKKUM
'Occasionally I (would) use English'

Features derived from language B which are found in the speech of A^0 speakers (speakers of A who are not in contact with any B speakers) can be considered fully integrated in A, as is the case with features in category (1). Such features can only *enter* the language from the other end of the chain of speakers, that is, through A^2 speakers/readers. Ideally, a complete description of any convergence situation would include, as far as possible, the level of penetration of each feature in terms of some such set of contexts as that suggested earlier. It might, for example, note that a feature X (a borrowed plural affix, say, or the loss of part of a grammatical rule) occurs invariably in certain contexts, variably in others, and not at all in a third set of contexts. It might also indicate, where relevant, whether borrowed affixes are used only with borrowed stems, or also with native stems; and any other sociolinguistic restrictions on a particular feature (e.g. class dialects) would also be noted. Needless to say, our sources rarely give any of this information.

4.32. "Native" vs "non-native" speakers

While the distinction between an individual's MT (mother tongue) and OT (other tongues) may be made easily enough in the early stages of contact, it often becomes difficult later on, if the groups become more intimately integrated with succeeding generations. At the stage when children typically learn both languages from an early age, the native/non-native distinction begins to become less important. It may still be possible, however, to distinguish between the *home language* (or *ingroup* language) and the *outside* (or *outgroup) language* of an individual or group. This is the case, for example, with the Saurashtri of Tamilnadu and the Kudumbi of Kerala, and presumably of Istro-Rumanian (at the time it was reported; see Cotéaunu 1957). At a still later stage, if intermarriage and interbreeding become common, even this distinction may disappear, and we can at most speak of *ethnic speakers* of a language: for example, many speakers who are bilingual in Dakhini Urdu and Telugu can be regarded as ethnic speakers of Telugu, in the sense of belonging to families which have retained Telugu names and who are probably descended from monolingual speakers of some form of Telugu. A similar situation may be found, for example, in North America, where some ethnic speakers of Spanish are totally fluent and literate in local varieties of English, but increasingly less competent in Spanish. At this stage, one may also speak of a language of *intragroup communication* as opposed to one of *intergroup communication*.

As long as this distinction at least remains, the type of theoretical constructs employed in **4.31** and in **4.4** may still be valid. However, a stage may be reached in which there is no longer any direct relationship discernible between ethnic and linguistic identity. At that point the investigation of linguistic convergence

becomes purely a matter of historical reconstruction, though the retention of ethnic names and customs may provide clues to the earlier identities and provenances of the groups involved.[10]

4.33. Inferences about range of contact

The examples in **4.31** show that the use of borrowed elements in a language varies according to the social context, in that they penetrate certain contexts more easily than others. It was noted, for example, that English words and expressions are avoided in formal speech and writing in the literary languages of India, though they are frequently used in conversation. The relevance of this point to linguistic archaeology is that in dealing with any particular ancient text, we need to recognize that we are looking at only a small segment of the convergence continuum. To locate our text in that continuum requires some knowledge (or at least some assumptions) about the linguistic habits and attitudes of the source culture, as well as an understanding of the interactions between language and social life in general. Just to mention a few obvious points, it can be generally assumed that religious and ritual texts are likely to belong to the monolingual end of the convergence continuum, that is, to avoid using expressions that are perceived as "foreign", since these texts relate to the core values of a society and resist intrusion by such elements.[11] This is certainly true for South Asian society, and probably for many others. An example from a very different part of the world is given by Stanley Newman, in a discussion of the distinction between ordinary and sacred language in Zuñi:

> ...obviously borrowed words, like **melika**, "Anglo–American", cannot be used in the kiva [the building used for religious purposes]. This prohibition against loan words is obviously not to be equated with the tradition of linguistic purism...It stems, rather, from the general Zuni injunction against bringing unregulated innovations into ceremonial situations: using a word like **melika**, as one informant expressed it, would be "like bringing a radio into the kiva."
>
> (Newman, in Hymes 1964: 398)

Other types of texts, such as commercial texts, might not be subject to this sort of linguistic inhibition – though of course, each cultural situation will have its own peculiarities. This point also has possible relevance to the analysis of borrowed material in the ancient Indo-Aryan ritual texts. See Chapter 3, Appendix, for a discussion of this case.

4.4. Models of linguistic diffusion

I discuss here four models of linguistic contact situations which are designed to account for the *routes* by which structural features diffuse from one language to another under different circumstances. Each of these models is depicted

diagrammatically, using the general scheme of Figure 4.1 plus the following additional conventions: (1) the underlined letter indicates the language most affected by convergence; (2) arrowheads indicate the direction of diffusion of features. While admitting that it is difficult to depict a dynamic process by a static representation, I offer these models for the purpose of making explicit some assumptions about the sociolinguistic processes involved in different types of convergence situations.

Figure 4.2(A) represents the 'native convergence' or 'adoption' model, which is applicable to the languages of small immigrant groups (such as "Pennsylvania

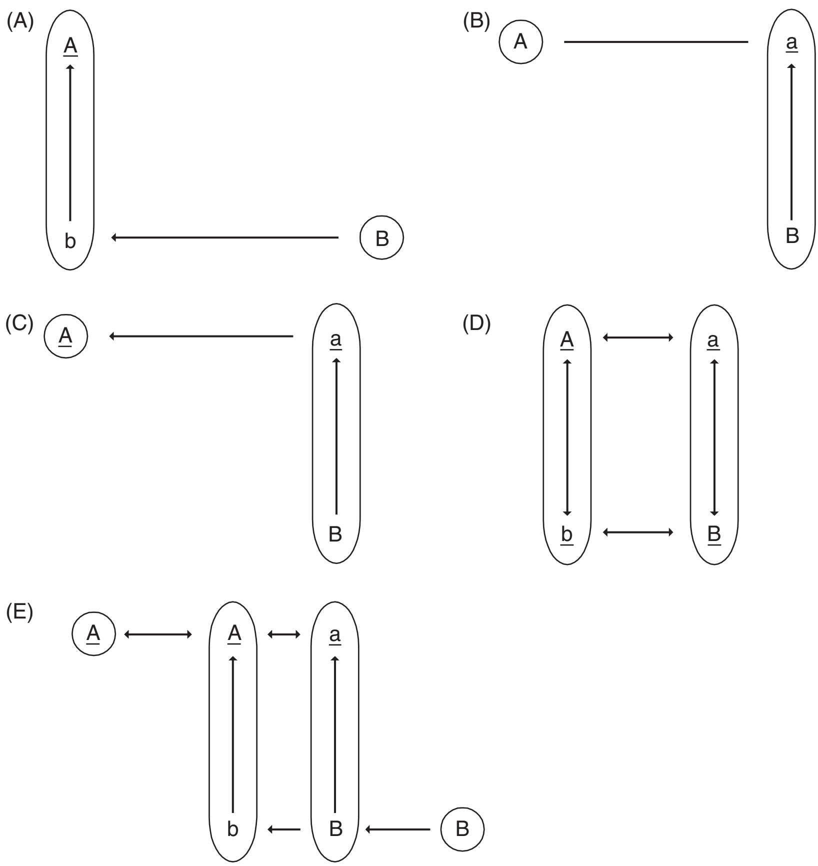


Figure 4.2 Models of linguistic diffusion. (A) Native convergence (adoption) model. Examples: Kudumbi, Saurashtri, Pennsylvania German, American Italian. (B) Substratum (carry-over) model. Examples: Indian English, Dakkhini Urdu, Italian English. (C) Bilingual majority model (Emeneau). Example: Brahui. (D) Mutual convergence model (Gumperz and Wilson). Examples: KuMarathi, KuKannada, KuUrdu. (D) Creolization-cum-convergence model (Southworth 1971). Example: Marathi.

Dutch," i.e. Pennsylvania German) which remain cohesive in their new habitat, retaining their language and many of their customs while becoming bilingual in the host language for economic, cultural, and/or political reasons. This model assumes the presence of two main groups, B^0/B^1 (e.g. speakers of English, mostly monolingual) and A^2 (e.g. speakers of German, mostly bilingual in English), and that the direction of diffusion is primarily from B to A, that is, convergence towards B (English) occurs in the *native* speech of the A^2 (German bilingual) speakers. The underlined letter **A** represents Pennsylvania German. An example of the working of this model would be the adoption of the English word **whatever** into Pennsylvania German in the form **was-ever** (see **1.22C**). In this model the possibility of monolingual B speakers is assumed, at least in the early stages of contact. In general it may be assumed that some members of some immigrant groups, such as women or the elderly, may have less contact with the host language than others. This model is also applicable to several languages of small immigrant groups in India, such as Saurashtri (Randle 1943; Pandit 1968, 1974; Ramasubramaniam 1968), Kudumbi (Thampuran 1981), and Kasargod Marathi (Ghatage 1970).

The second model, the 'substratum' or 'carry-over' model, applies generally to situations in which members of a group, or several groups of different language backgrounds, acquire the language of a smaller dominant group, usually for economic and/or political reasons. Indian English (see discussion in **4.22**) and Hawaiian English are examples. This model assumes that the B^2 speakers (e.g. Tamil-English bilinguals) create a new version of A (e.g. English) by *carrying over* features from their own language, and that the existence of this hybridized form of A has no appreciable effect on the speech of the majority of A^1 (e.g. monolingual English) speakers. (If there are any **A–b** speakers (e.g. native speakers of English bilingual in Tamil or Malayalam), their role in the development of the new form of A is assumed to be unimportant. An example for this model: features of English and one or more Indian languages are combined to produce Indian English expressions like those illustrated in **4.31**(A). This model can also apply to the version of the host language spoken by immigrant groups, though these versions (such as American Italian or Pennsylvania English) are often short-lived.

Note that each of these models is designed to account primarily for the history of a single language. While clearly both of these models can apply to situations like that of the Pennsylvania Germans, the first is meant to represent the changes undergone by their native German, while the second represents their version of English.

Figure 4.2(C) represents the "bilingual majority" model, named after a proposal by Emeneau regarding the diffusion of structural features from Balochi (Iranian) into Brahui (Dravidian):

> At one time in the history of the Brahui Confederacy there must have been more non-native speakers of Brahui whose mother tongue was Balochi, and descendants of such speakers, than there were speakers who had learned the language from native speakers. It was this *bilingual*

> *majority* [my italics] who handed on to later generations their version of Brahui, a version which in many features was essentially a calque of Balochi clothed for the most part in Brahui forms.
>
> (Emeneau 1962b: 60, reprinted with permission of the publisher from M. B. Emeneau, Brahui and Dravidian Comparative Grammar, Berkeley and Los Angeles: University of California Press, 1962)

This model differs from the second model in that, since the B^2 group carrying over features from its native speech (B = Balochi) to its non-native speech (**a** = Brahui) is the majority, the new version of **A** (Brahui) is also adopted by the minority A^1 speakers and their descendants.[12] This model assumes the existence of monolingual speakers of both A and B, at least in the early stages, though Emeneau's (1962a) discussion of the Brahui situation suggests that in recent times, most speakers of Brahui and Balochi have been bilingual (but see the previous comment).

The fourth model, called here the 'mutual convergence' model, represents a situation found in a village in the Maharashtra–Mysore border region (Gumperz and Wilson 1971). Gumperz and Wilson found that of the grammatical changes observed in the local languages (as compared with varieties of those same languages spoken elsewhere), ten changes were toward Kannada and ten toward Marathi. This symmetry of mutual convergence, represented in the Figure 4.2(D), is explained by Gumperz and Wilson in terms of a balance between the economic dominance of Kannada in village life (the landowners being mainly ethnic speakers of Kannada) and the political dominance of Marathi in the surrounding area – Marathi being the official state language and the official medium of education in the local schools. Generally speaking, this model seems applicable to situations where the status of two languages is fairly equal or in some way equivalent, and where bilingualism is widespread in both groups. Such situations are perhaps most likely to appear in border areas. Note that, as opposed to model 4, models 1–3 all involve asymmetrical relations between languages such that language A is learned by speakers of B for reasons of prestige, cultural or economic dominance, or the like. This last model does not deny the presence of monolingual speakers of A and B, but assumes that they are not primary actors in the public life of the locality.

Figure 4.2(E) presents a model implied by Southworth's (1971) discussion of the problem of the "Dravidianization" of Marathi, assuming that members of the B^2 group (Dravidian speakers using Dravidianized varieties of Indo-Aryan) functioned as intermediaries between the A^1 group (monolingual Indo-Aryan speakers) and the B^1 group (monolingual speakers of Dravidian). The A^2 group (native IA speakers bilingual in Dravidian) may have developed at a later stage. According to Southworth's proposal, the varieties of Indo-Aryan spoken by these various groups ultimately converged into a single continuum, with Dravidian being later abandoned in many areas (Southworth 1971). With a large indigenous population of monolingual B^1 speakers, compared to a much smaller group of immigrant A^1 speakers, the diffusion of structural features might be largely

unidirectional. This model may be regarded as a more elaborate version of that shown in Figure 4.2(C), the bilingual majority model.[13]

Either model C or model E would probably also be appropriate for the situation in which the early speakers of Indo-Aryan came in contact, over a period of several centuries, with speakers of indigenous languages ("Indus," Munda, AA, Dravidian, etc. see **3.2**). Model A appears inappropriate here because Indo-Aryan is assumed to have been a prestigious language for political or economic reasons – mainly because Indo-Aryan languages subsequently became dominant in these areas. Model B is ruled out here because it assumes that there are A^1 speakers who are unaffected by convergence. Model D might be applicable here, though we do not know enough about the impact of OIA on the indigenous languages in the early period of contact.

A decision about which of these models to invoke in explaining a particular case in the past will depend on the direction of diffusion of structural features. Model A involves diffusion from the host language to the native language of an immigrant group (*adoption* of features), whereas model B involves interference in the immigrants' version of the host language (*carryover* of features). Models C and E are similar to B, with the difference that in these cases there is no surviving *unconverged* variety of language A. Model D involves structural interference, in more or less equal proportions, in both directions. Where the historical record provides adequate data on these matters, we can hope to be able to reconstruct at least some notion of the socio-historical circumstances in which the earlier contact took place.

4.5. Summary

This chapter seeks to provide an empirical basis for drawing inferences from data on linguistic convergence (linguistic elements diffused from one language to another) regarding the socio-cultural contact situations in which the convergence took place. Three types of inference are discussed: *intensity* of contact (the extent of social integration of the groups involved, **4.2**), *range* of contact (the proportion of each group involved in the contact, **4.3**), and the *dynamics* of contact and the relationships among the groups involved (**4.4**). Examples are given from modern and ancient languages of South Asia and elsewhere.

Following are the major findings of this investigation relating to the intensity of contact:

(1) percentages of borrowed content words reflect primarily the *duration of contact*: thus, 500 years of Turkish rule in the Balkans produced higher levels of lexical borrowing than 200 years of British rule in South Asia;
(2) structural convergence (borrowing of grammatical or phonological *rules*), or borrowing of function words or inflectional affixes, or high levels of core (non-cultural) borrowing, reflect *symbiotic* relationships between ethnolinguistic groups in which the borrowing group identifies itself as part of the same society as the other group: thus the length of time for Turkish in the

Balkans is comparable to the period of Saurashtri contact with Tamil, but the latter case shows much greater grammatical convergence;

(3) extreme levels of the types of convergence mentioned under (2) imply at least a partial loss of the separate ethnolinguistic identity of the borrowing group;
(4) while intermarriage and intrafamilial bilingualism are presumably important social mechanisms promoting linguistic convergence, they are neither sufficient nor necessary explanations from a cross-cultural point of view;
(5) though cultural identification between groups is a *concomitant* feature of high levels of linguistic convergence, it is probably not an adequate *explanation* of it; *economic interdependence* would seem to be a more potent, and more general, factor.

Regarding the range or extent of contact, it is important to recognize that borrowed material in a language (whether lexical or structural elements) may not be available to all speakers of the language, and may not be used in all contexts. In particular, religious texts tend to be governed by notions of cultural or linguistic "purity," and their creators may avoid conscious use of borrowed material. Thus, if we find even a few borrowed words in a society's ritual texts, we would conclude that other parts of the *convergence continuum* (**4.31**) would probably show a higher level of borrowed elements.

Section **4.4** proposes five models for such situations, which differ according to the quantities and types of elements diffused between two languages, as well as the direction of diffusion:

(A) the native-convergence (adoption) model, applicable to small immigrant groups who modify their home language by adopting elements of the host language (e.g. Pennsylvania Dutch, Saurashtri);
(B) the substratum (carry-over) model, applicable to the version of a host language, or a colonists' language, spoken by those (immigrant or colonized groups) who carry over elements of their home language into it (Indian English, Hawaiian English);
(C) the bilingual majority model, referring to the situation of Brahui, in which the language of an originally dominant minority has accepted linguistic diffusion from local languages through widespread bilingualism of both A and B speakers;
(D) the mutual convergence model, applicable to situations in which two languages of comparable prestige show similar levels of convergence with each other (Gumperz and Wilson's Kupwar Marathi and Kupwar Kannada);
(E) the creolization-cum-convergence model, a more elaborated version of model C which posits two bilingual groups as intermediaries between the monolingual extremes.

The model most applicable to Old Indo-Aryan, in its early contact with Dravidian, appears to be either C or E.

Notes

* This chapter appeared in an earlier form as Southworth 1990a. This revised version is published here with permission of the Institute of Linguistics, Faculty of Philosophy of the University of Zagreb. I am grateful to Professor Damir Kalogjera for useful discussions on this and related topics.

1 For example, George Trager (1939) notes that the word for 'cottonwood' has become a generic word for 'tree' in a number of languages of the southwestern United States.

2 Some cases may be difficult to decide, and some cases may involve both types of change, as in Heath's example of the borrowing of a derivative suffix which facilitated the later borrowing of verbs of a particular class (1981: 364). The borrowing of the Proto-Dravidian quotative particle ***viṯṯi** into OIA (**3.23**, Chapter 3 Appendix Item 1), as well as the similar borrowing of a Malayalam quotative marker in Kasargod Marathi (**3.32**), appears to be a similar case.

3 The distinction between substantives and other words may not be easily applicable to many languages. In Hawaiian, for example, the word **aloha** can translate as a noun ('love, affection, compassion, mercy, sympathy…'), as in **me ke aloha o Kawena** 'with the love (or greeting) of Kawena'; as a verbal: **ua aloha A i B** 'A loves B'; as a modifier: **mea aloha** 'loved one' (lit. "thing/person [of] aloha"), **hoa aloha** 'friend' ("companion [of] aloha") – compare **hoa kula** 'schoolfellow' ("companion [of] school"); and as an interjection: **aloha 'oe**' 'greetings/love to you'(Examples from Pukui and Elbert 1986, Snakenberg 1988). Pukui and Elbert list **aloha** as *nvt, nvs*, that is noun–verb transitive and noun–verb stative. (Stative verbs can generally function as modifiers.) Even for such languages, the distinction between content words and grammatical words will still be valid, by and large; whether the former category can be meaningfully subdivided (e.g. concrete nouns vs other content words) can only be found out by further empirical investigation.

4 Damir Kalogjera (1991–92) has described a situation in southern Dalmatia which is at variance with the generalizations made here. The local version of Croatian contains many Italian lexical items (including not only substantives but also adjectivals, conjunctions, and discourse features like **dunkve** = Italian **dunque** 'Well then…') as well as phrases and syntactic calques, though on the whole the local people are not able to carry out meaningful conversations in Italian – even if they can occasionally bluff an Italian speaker for a short period of time. Kalogjera explains this situation as a residue of an earlier period (fifty or more years earlier) in which Croatian–Italian bilingualism was general in the area for both native Croatian and native Italian speakers, at a time when Italian was regarded as the more prestigious of the two languages.

5 For example, **tarah** 'way, type, manner', **lekin** 'but', **hazār** '1,000', **bād** 'after', **mālūm** 'apparent, known', **ādmī** 'man, person', **yād** 'memory', **kāfī** 'enough', **giraftār** 'arrested', **sarkār** 'government' (these are the most common colloquial words for these meanings).

6 The structural influence of French on English can be clearly demonstrated in two areas, both of which might be considered to lie in an intermediate position between the lexical and the syntactic realm: (1) the introduction of new affixes from French (or Latin) such as **re-**, **dis-**, **trans-**, **counter-**, **-able**, **-ment**, **-(a)tion**, **-ity**, along with the atrophy of older affixes such as **for-** (**forbear**), **to-** (**tobrake** 'break up', **G. zerbrechen**), **with-**(**withstand** 'stand against'), **-lock** (**wedlock**), **-red** (**hatred, kindred**), **-dom** (**kingdom, freedom**), which eventually lost their productivity and were either lost or became restricted to individual lexical items (Baugh 1957: 218–21), a process described in detail in Dalton-Puffer 1996; (2) the adoption of thousands of phrasal calques, such as **put in prison** (**F. mettre en prison**) and **beforehand** (**F. avant la main**), which led to a style of English which Prins (1952: 302) regarded as mirroring the thought processes of French speakers. However, it is difficult to make the case that

syntactic changes of the type described in this chapter resulted from contact with French, even though a three-way comparison of English, French, and German shows English to resemble French more closely in some features, such as word order in sentences. (Masica's study of linguistic areas indicates, for example, that English falls almost exactly midway between French and German in its frequency of use of participial constructions; see Masica 1976: 134.) Baugh and Cable (1993: 154ff.) argue that most of the grammatical changes which took place in Middle English were *not* the result of contact with French. Fischer, writing on English syntax in the *Cambridge History of English* (Vol. II), does not discuss French influence except to mention that the development of the periphrastic genitive (**the pen of my aunt** as opposed to **my aunt's pen**) "may have been helped along by the parallel French construction" (1992: 226). Equally important, the influence of French on English was initially, and continued to be, from the top down – similar to the situation of Persian in South Asia – while, unlike that situation, it was exercised more through the written language than through spoken contact.

7 Furthermore, the Kudumbis no longer celebrate their traditional festivals separately from the local celebrations, and Thampuran (1981: 7) reports that they have petitioned the local government to recognize them as a depressed local caste group (in order to obtain concessions in employment, admission to schools, etc.).

8 For example, in recent years Korean immigrants have moved into the fruit-and-vegetable business in many North American cities, just as Italian immigrants in earlier generations moved into building construction, shoe repair, and barbering.

9 E speakers who use these words are not necessarily aware that they are of Indian origin; S speakers who use them in English are not necessarily aware that they are different from the words in the next category.

10 Brahui may represent this stage, given Emeneau's description of the complicated linguistic situation among the Brahui (see Emeneau 1962a: 49–50). In many parts of the UK, Irish, Welsh, and Scottish surnames are often the only indications that people's ancestors spoke some form of Gaelic.

11 This would obviously *not* be the case when the religious tradition itself is linked to another language, for example the use of Latin and Greek in the European Christian traditions.

12 Note Emeneau's further comment: "Whether in the Brahui–Balochi situation it is necessary or just to speak in terms of 'upper' or 'lower', is still what we do not know. Nor in the sorry state of our knowledge of Balochi, can we be sure that it too has not been Brahuized in some structural features." (1962a: 60) Thus it may be that Gumperz and Wilson's "mutual convergence" model is more appropriate here.

13 Despite the use of the term 'creolization' in the name of this model, it is not necessary to assume that pidginized versions of either language resulted from this process, except possibly as transitional variants. Thus I would no longer maintain that Marathi underwent a process of pidginization, as posited in Southworth 1971.

5

THE GRIERSON HYPOTHESIS REVISITED

Subgroups of Indo-Aryan

Contents

5.0. Introduction: the problem of subgrouping in Indo-Aryan

In **1.22B**, the question of linguistic subgrouping was discussed in general terms. The present chapter addresses the problem of subgrouping in NIA, using Masica's summary of the situation as a starting point (1991: app. II, pp. 446–62). Evidence will be presented in **5.1** and **5.2** for a number of shared innovations in a group of languages traditionally known as the "outer group" of Indo-Aryan. Chapter 6 will examine the implications of this linguistic evidence in the light of relevant archaeological and historical knowledge.

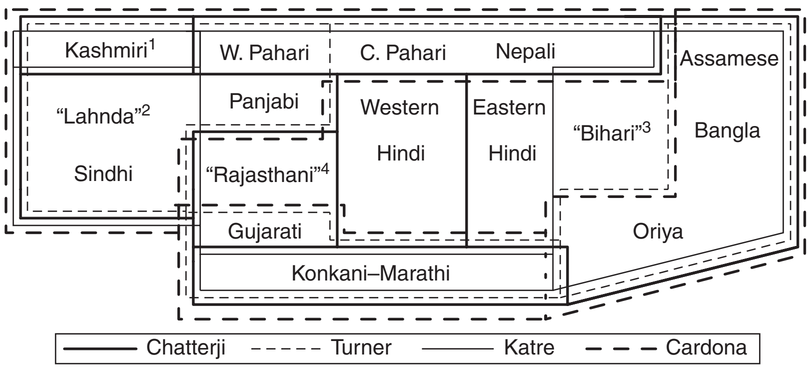


Figure 5.1 Various subgroupings of NIA.

Notes

1 Kashmiri is often considered to belong to a "Dardic" branch of Indo-Aryan, though this seems to be largely a regional grouping with some locally diffused shared features, rather than a genuine genetic subgroup.
2 "Lahnda" or "Lahndi" refers to the western varieties of Panjabi, now entirely in Pakistan.
3 "Bihari" is a language construct of Grierson's, comprising Maithili, Bhojpuri, and Magahi.
4 "Rajasthani" is also a Grierson construct, which includes the dialects located between Hindi and Gujarati.
5 Not included here: (A) Rumany/Rumani, the languages of the Gypsies (probably connected with the Dumaki language of Hunza); (B) Dardic (see note1); (C) Nuristani, a group of languages in the extreme northwest of the subcontinent; (D) Sinhala, the language of Srilanka, and the closely related Maldivian of the Maldives.

Masica (1991) discusses the history of NIA subgrouping, and notes that there is little agreement among scholars on the locations of subgroup boundaries. Figure 2.1 shows the approximate locations of the major Indo-Aryan languages. Figure 5.1 is a schematic diagram combining the proposed subgroupings of S. K. Chatterji (1970), R. L. Turner (1975), S. M. Katre (1968), and G. Cardona (1974), as presented in Masica 1991: 452–5. The reason for this lack of agreement is the frequent overlapping of structural features in dialect space: thus for example Cardona, who puts Gujarati and Marathi–Konkani in the same "West and Southwest" subgroup, notes that these languages share the mergers of **i** with **ī** and **u** with **ū** (1974: 448).[1] On the other hand, Marathi–Konkani differs from Gujarati in having an **-l-** future (M **karil** 'he will do': G **karśe**) and a set of alveolar affricates **t^s d^z**, which contrast with the palatals č ǰ), among other features, while Gujarati's phonemically distinct open vowels **ɛ** and **ɔ**, its **s-** future[2] (see previous example), and other features link it with varieties of Rajasthani and Hindi. Masica's view of the situation is that criteria exist which can justify all the subgroupings which have been proposed, but since these criteria conflict, the question of whether or not to group any two languages can only be decided by "giving priority to some criteria over others" (1991: 456).

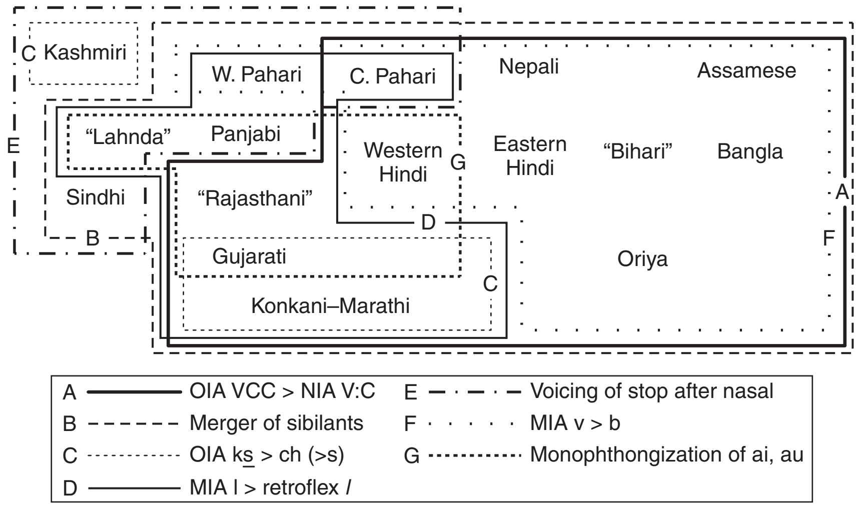


Figure 5.2 Some phonological isoglosses in NIA (after Masica 1991: 459, reproduced with permission from C. A. Masica, *The Indo-Aryan Languages*, Cambridge: Cambridge University Press, 2003).

Figure 5.2, which is similar to Masica's figure II.10 (1991: 459), indicates the nature of the problem. As Masica points out, it makes no difference if morphological criteria are used instead of phonological features, since these also show overlap: for example, Marathi–Konkani shares the past suffix **-l-** with the Eastern languages (see below), and the future suffix **-l/n-** with Rajasthani and Nepali, while future forms in **-s/h-** are found in a broad band extending from Sindhi and "Lahnda"[3] to eastern Hindi, cutting across the area of the **-l/n-** futures (see Masica 1991: 288ff. and the diagram on p. 290). Masica concludes as follows:

> Perhaps a wiser course would be to recognize a number of *overlapping genetic zones*, each defined by specific criteria…the mixed dialectal ancestry of most NIA languages…would at least be clear. We might therefore be well-advised to give up as vain the quest for a final and "correct" NIA historical taxonomy, which no amount of tinkering can achieve, and concentrate instead on working out the history of various features, letting such feature-specific historical groupings emerge as they may, with their overall non-coincidence as testimonial to the complexity of the situation.
>
> (1991: 460, reprinted with permission from C. A. Masica, *The Indo-Aryan Languages*, Cambridge: Cambridge University Press, 1991)

It is not entirely clear what the phrase "overlapping genetic zones" means. On the one hand, overlapping zones are the norm in any dialect continuum, as has been shown repeatedly by the work of dialect geographers in France, Germany, and Britain,[4] and it is generally recognized that in long-settled areas – for example under sedentary agriculture – linguistic innovations diffuse in wave-like fashion until they meet a natural or socio-political boundary (see **1.22B**, **1.31**). Since such boundaries change from time to time, and since the relative importance of particular urban centers and other centers of linguistic innovation waxes and wanes, the result over time is a network of overlapping isoglosses such as that shown in Figure 5.2. Historical linguists would expect this to be the case in any large populated area, even when the starting point is assumed to be a region of relatively uniform speech. On the other hand, to call such overlapping zones "genetic" seems to imply the possibility of identifying *original* zones of some kind, precisely the possibility that Masica appears to reject.

In fact, most of the isoglosses shown in Figure 5.2 do not correspond to any of the subgroup boundaries shown in Figure 5.1.[5] Thus the scholars who have dealt with this question all agree, at least implicitly, that most of these innovations are not diagnostic for subgrouping. This clearly implies that if there are any valid *genetic* subgroupings, they must *predate* most of these innovations (see **1.22B**). Is there any way to reconstruct original subgroupings? Masica's position seems to be that the overlapping pattern we observe makes it impossible to define subgroups without arbitrarily selecting some criteria over others. This chapter takes a different position: if there is evidence to establish the *relative chronology* of innovations, it then becomes possible to establish the stratigraphy of dialect divergence (see discussion in **1.22B**). In the absence of dateable textual evidence, or other documentation of the chronology of linguistic changes, the *distribution* of innovations is the only possible source of such evidence.[6] When a number of innovations have similar boundaries which do not correspond to any current natural or political divisions, such boundaries may well be a residue of earlier sociolinguistic divisions: see further discussion in **5.3**.

Given a group of related languages **A**, **B**, . . . **N**, two or more of them can be said to belong to a single subgroup only if they show evidence of undergoing one or more *exclusively shared innovations*. All three terms are crucial: the feature in question must be the result of an innovation, since shared *retentions* in two languages can easily be independent of each other. It must be *shared by all* languages in the putative subgroup, and must be *exclusive* to those languages. Thus for example, the changes which transformed Proto-Indo-European short **e** and short **o** to **a**, and long ē and ō to ā are shared by all Indo-Aryan and Iranian languages, and by no others, and therefore constitute defining features of the Indo-Iranian subgroup of Indo-European (Buck 1933: 78–9). On the other hand, the change **s** → **h** which Grierson claimed as a characteristic of the outer group (see later) is primarily a western phenomenon, and is not shared by the eastern members of the proposed group. The retention of the three Indo-European gender categories (masculine, feminine, neuter) in Marathi, Gujarati, Sindhi, and other western and

northwestern NIA languages, which has been occasionally adduced as evidence of common history, cannot be considered relevant in this context as it is not an innovation.

5.1. The Grierson hypothesis

Sir George Grierson, the compiler of the *Linguistic Survey of India* (hereafter *LSI*), proposed a genetic division of the modern Indo-Aryan (NIA) languages into three major subgroupings, a proposal which has since been referred to as the "Grierson hypothesis," or "Hoernle–Grierson hypothesis" (since a similar division had earlier been suggested by Rudolf Hoernle).[7] On the basis of the distribution of certain features of the modern languages, Grierson spoke of a basic division between "inner," "intermediate," and "outer" sub-branches of Indo-Aryan languages, as depicted in Figure 5.3.

While Grierson altered some details of membership in the various subgroups,[8] the important point for the present purpose is that he retained the belief that the eastern languages (Bangla, Ahom/Assamese, Oriya, and "Bihari") were more closely related to Marathi and the northwestern languages ("Lahnda" and Sindhi) than any of these were to the central group (western Hindi, Pahari, Nepali).

5.11. Grierson's view and Chatterji's critique

Grierson's evidence for his hypothesis was presented sketchily in the introduction to the *LSI* (1.1.116–18), and in greater detail in his article "On the modern

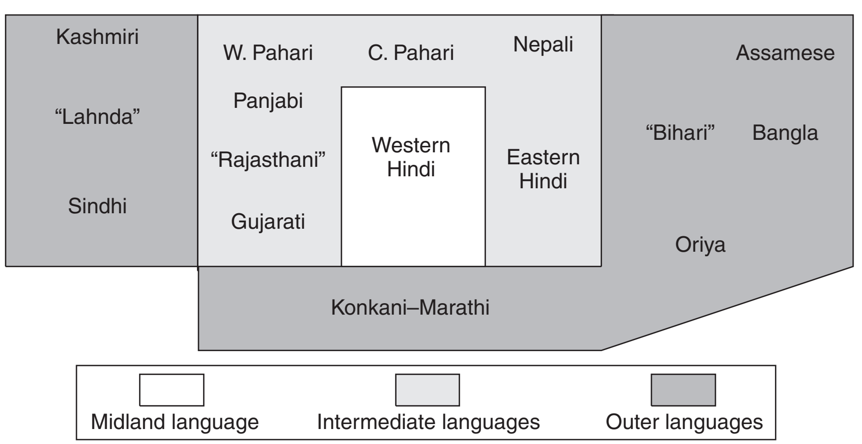


Figure 5.3 Grierson's view of Indo-Aryan (after Masica 1991: 453, reproduced with permission from C. A. Masica, *The Indo-Aryan Languages*, Cambridge: Cambridge University Press, 2003).

Indo-Aryan vernaculars" (Grierson 1931–33). S. K. Chatterji (1970: 32–33, 150–69) examined Grierson's evidence point by point, arguing that for the most part it fails to make the case for an inner–outer division, either because (a) the geographical distribution was misstated by Grierson, or (b) because the phenomena involved are very sporadic, or (c) because the shared features represent common retentions from an earlier period, or (d) because the differences are of very recent origin. For example, the change of intervocalic **s** to **h** is primarily a western phenomenon, though in the numerals it is found throughout NIA; the retention of final vowels in the west (Kashmiri and Sindhi) and some of the eastern languages (e.g. Maithili) is no evidence of a close connection since shared retentions can be independent of each other.

5.12. Grierson's evidence: past forms in -l-

Perhaps the most important evidence adduced by Grierson was the past form in **-l-**, which occurs both (A) as a past verbal form and (B) as an attributive verbal adjective; it is found in one or both forms in the languages from southern Pakistan (Sindhi) to the west coast (Gujarati), the Deccan (Marathi), across central India south of the Vindhya complex to the east coast (Oriya), the lower Ganga plains (the 'Bihari' languages Bhojpuri, Maithili, Magahi), the Ganga–Brahmaputra delta (Bangla) and Assam (Ahom), as depicted in Grierson's map in LSI Volume 1, Part 1, p. 140.

*5.12A. Past indicative in -**l**-*

The **-l-** suffix is the regular past or perfective marker in a continuous area stretching from Assam in the east to Gujarat in the west, as well as in some "Dardic" languages.[9]

Examples[10] (all citations are of the form Verb + Past(+ Concord), and translate as '(he) went'):

Brokpa (Shina)	gā-**l**-o (LSI 8.2.201)
Gujarati	ga-y-o, ga-(y)**el**-o[11]
Marathi	ge-**l**-ā
Konkani	ge-**l(l)**-o
Halbi	ge-**l**-ā (LSI 7.363)[12]
Oriya	ga-**l**-ā (LSI 1.2.331)
Bhojpuri	ga-**il**
Maithili	ge-**l**
Magahi	ge-**l** (LSI 1.2.331)
Bangla	gæ-**l**-o
Assamese	ga-**l** (LSI 1.2.331)

The inner languages, on the other hand, show forms derived from the OIA past (passive) participle in **-t-**:[13]

Hindi–Urdu	ga**-y-ā** 'gone, he went' (← MIA ga(y)a-, OIA gata-)[14]
Panjabi	ge-ā
Sindhi	hal**-i-**o (from hal-aṇu 'go', Trumpp 1970: 317)
Awadhi	gā
Marwari	ga**-y-**o

*5.12B. Past/perfective participle in **-l-***

A related form, the adjectival participle in **-al/il/el-**, also occurs in Sindhi, Maiyã (a northwestern language), Gujarati, Bhili, Marathi–Konkani, Khandeshi (Ahirani), some varieties of Halbi (see note 12), and the eastern languages:

Maiyã	kuṭ**-el** as 'was beaten' (LSI 8.2.525)
Sindhi	dhot**-al-**a kapṛa 'washed clothes' (C. J. Daswani, p.c.)
Gujarati	lakh**-el-**o patra 'the written letter'
Bhili	pik-n-**el** 'ripe', kun-n-**el** 'rotten' (Kulkarni 1976: 56)[15]
Old Marathi	tuṭ**-al-**i sāuli 'broken shadow'
Marathi	kel**-el-**e kām 'the done work'[16]
Khandeshi	kar**-el** kām 'the done work' (Chitnis 1964: 77)
Konkani	pik**-all-**o ambe 'ripe(ned) mangoes'
Early Maithili	nahā**-il-**i gorī 'bathing [lit. bathed] fair (woman)' (ODBL 958)
Bhojpuri	sūt**-al** ghodā 'sleeping horse' (Shukla 1981: 119)
Early Oriya	arj**-il-ā** bibh**ū**ti 'earned merits' (ODBL 958)
Old Bangla	ge**-l-**a jāma 'gone (i.e. previous) birth' (ODBL 956)
Middle Bangla	pāk**-il-**a drākṣā 'ripe(ned) grapes' (ODBL 957)
Bangla	ge**-l-**o bɔchɔr 'last (i.e. gone) year' (ODBL 956)
Bangla (dialectal)	kɔra**-il** kam 'completed work' (ODBL 956)[17]

These **-l-** forms are generally agreed to have originated from a MIA adjective-forming suffix **-illa/alla/ulla-** (see Bloch 1965: 267–8; ODBL 940). It is clear that this suffix was added to the original *past* stems of verbs, that is, the forms containing the OIA past participle in **-t-**, for example Marathi **gelā** 'gone' from ***gayalla** from ***ga-t-alla, kelā** 'made' ← ***kayalla** ← **kṛ-t-alla** (Master 1964: 123, ODBL 941), though in many cases the past stems have been replaced by the present stems through analogical change (e.g. Gujarati **kar-elo** 'did/done'). This **-ll-** suffix appears to be the same suffix which is used to create adjectives from postpositions, adverbs, and other forms, as in Marathi **ātlā** 'inside, internal' (← ***antr-alla**, cf. **āt** 'in, inside' ← OIA **āntra-**), Hindi–Urdu **aglā** 'next' (**āge**

'forward', cf. OIA **agra**), **pichlā** 'previous' (**pīche** 'back'); see ODBL 167 for further examples. Though the suffix occurs throughout NIA, its specific use as a past verbal form in the outer languages is clearly an *innovation vis-à-vis* the languages of the inner group, in which the past/perfective forms lack the **-l-** suffix.

Other possible origins for the **-l-** pasts have been mentioned, at least for some of these languages. In the European Gypsy languages, **-l-** is apparently the regular reflex of intervocalic **-t-**: **gili** 'song' (OIA **gīta**), **phral** 'brother' (OIA **bhrātā**), **perel** 'falls' (MIA **paḍai**, OIA **patati/paṭati**), **śel** '100' (OIA **śatam**), **śel** 'cold' (OIA **śītaḥ**) (Turner 1975: 264–5). Past forms in Sinhala containing **l** are derived by Geiger from composite verbs containing the auxiliary verb **lanu** 'put, place' (Geiger 1938: 137, 159). Chatterji (ODBL 943–4) mentions possible OIA origins for the MIA **-ll-** suffix; he also notes (and rejects) the suggestion of a link between the Indo-Aryan **-l-** pasts and the pasts in **-l-** found in Slavic and elsewhere in Indo-European.[18] (The fact that the **-l-**is added to the earlier *past participle* in **-t-** would seem to support Chatterji's view.)

5.13. Diagnostic status of the -l- past

Chatterji dismisses the **-l-** past as a characteristic of the outer group in the following words:

> The adjectival "l" affix is an Indo-European inheritance in NIA, and occurs in the Midland speech as well: only, in the Eastern languages and in Marāṭhī, it forms the basis of the past tense, and in Gujarātī and Sindhī, it is regularly employed in a passive participle form.
>
> (ODBL 167)

The Indo-European origin of the **-l-** suffix notwithstanding, as a past form it is an innovation of the outer languages, as shown by the forms cited earlier. This appears to be a paradigmatic case of a common innovation, which has not been acknowledged as such by Chatterji and other scholars, perhaps because of the dilemma which such recognition might create (see **5.3** and **6.4**).

Was this an *exclusively* shared innovation in the outer languages? Bloch (1965: 270–1) notes that **-l-** forms are found in some northwestern languages, in European Romany, in Old Hindi (e.g. Kabir **pucchala, bādhala**), and in "hindi vulgaire" (from which he cites **gayalā** 'gone', **beclā** 'sold'. As noted earlier the **-l-** forms in Romany are the regular reflex of OIA **-t-**. This may also be the case for some of the northwestern languages; alternatively, it may be possible to include some of them in the outer group, as Grierson does – though there is not sufficient evidence at present to include them. It is probable that the **-l-** forms in Kabir are Bhojpuri or Maithili forms (see ODBL 99ff.), and the same is true for the "vulgar Hindi" forms quoted by Bloch. The LSI shows no trace of them, Bloch cites no sources for them, and no such forms are found in the Hindi grammars.[19]

Thus the available evidence points to the **-l-** past (apart from some northwestern languages) as an exclusive feature of the outer group languages.

Can it be established that the **-l-** past was indeed a *shared* innovation in the languages in which it appears? More specifically, can it be shown that it was shared by (some of) the eastern languages and (some of) the western languages? Answering this question requires us to distinguish three distinct innovations, in the following historical sequence:

(1) **Vp+{past} → Vp+{past}-l/__-C+Noun** (The "pleonastic" **l-**suffix is optionally added to the past base of a verb in the attributive position before a noun.) For example: an OIA **upaviṣṭ-o manuṣyaḥ** 'seated man' would become Proto-Outer ***upaviṭṭha-ll-o manussa** (cf. OB **baiṭhelo mānuṣ**, ODBL 946–7); OIA **pakvo āmraḥ** 'ripe(ned) mango' would become Proto-Outer ***pakk-all-o ām(b)a** (cf. Konkani **pikallo āmbe** 'ripened mangoes' cited above).
The reasons for assuming that this was the first phase of the innovation are as follows:

- (a) the **l-**forms appear to have been originally adjective-forming suffixes in OIA[20] and MIA (Pischel 1965: 406), and occur widely as such from the MIA period on in both inner and outer languages (see ODBL 167, 941; Bloch 1919: 256 for Marathi; Cardona 1965: 135 for Gujarati);
- (b) **l-**forms in this position (i.e. in a past participial form which modifies a following noun) have the widest distribution of the **l-**suffixes in the modern outer languages, modern standard Bangla being the main exception (see below).

The following two changes extended the distribution of the **l-** past into two additional environments:

(2) **Vp+{past} → Vp+{past}-l/__-C + Aux** (The **l-** suffix is optionally added to the past base of a verb when followed by an Aux.) For example, an OIA **gato (b)havasi** 'thou art gone' → Proto-Outer ***gata-ll-o (b)havasi**; cf. OM **gelāsi** (contracted from ***gelo āhāsi** ← presumed ***gatallo ābhavāsi**), NM **gelās**.[21]

(3) **Vp+{past}→Vp+{past}-l/__-C###** (The **-l-** suffix is optionally added to the past stem of a verb in clause-final position: for example OIA **so gataḥ** 'he went' would appear as Proto-Outer ***so gata-ll-aḥ)** (= M **to gelā**, B **śe gælo**).

Later, most of these cases of **-l-** ceased to be optional in all positions, but at different times in different parts of the outer group. These innovations appear to have begun in the west, perhaps specifically in pre-Marathi–Konkani, for several reasons. First, innovations (1) and (2) went to completion before the earliest

Marathi records, while in the east the **-l-** is variable in the early texts in all three environments (ODBL 944–59).[22] (In Gujarati, the **-l-** is invariable in attributive position, and remains variable in predicate position, as noted ealier.) Second, the Prakrit forms which have been quoted as the progenitors of the **-l-** past are apparently from western Prakrits, and have parallel forms in the modern western languages: Maharashtri **miliellau, milielaya** 'encountered' (cf. NM **miḷāla), jimiallaya** 'dined' (= OM **jevile**, M **jevle**, v. Tulpule 1960: 113); Jaina Prakrit **āgaelliyā** 'arrived' (Master 1964: 123; cf. NM **ālā**, Ko **āylo** 'came' – but note Oriya **āy-lo** 'came').

Thus these innovations, even the first and most widespread of them, *cannot* count as shared innovations in the strict sense (usually thought of as taking place simultaneously in a single uniform dialect), since the evidence appears to show that they spread over a period of time from one area to another. Thus we may be dealing here with a phenomenon similar to the lexical diffusion of sound change (see **1.21C**), in which a change moves gradually both in geographical space and in different linguistic environments. See further discussion in **5.3**.

The area in which the **-l-** forms occur is geographically unbroken. There is other evidence of linguistic links between east and west, some of which will be presented in **5.2**; the point will be discussed further in **5.3**. Here we may note the existence of certain shared irregular past forms, which occur in Marathi–Konkani and one or more of the eastern languages:

Konkani	ās-i-lo: Bangla āch-i-lo 'was'[23]
Marathi	ge-lā: Bangla gæ-lo 'went'
Konkani	āy-lo: Oriya āy-lo 'came'
Marathi	ghe- 'take', past ghet-lā: Inscr. Oriya ghet-alā (Misra 1975: 138)

Given this evidence, it would be normal to assume that these are related rather than independent changes. There remain questions about chronology, which will be discussed in **6.4**.

5.2. Additional evidence

Clearly, if the situation is as Grierson surmised, we should expect to find some additional linguistic data to support the inner–outer hypothesis. Thus it seems necessary to ask whether there might be additional evidence which Grierson overlooked, that is, innovations exclusively shared within either the inner or outer group. In the case of the outer languages, any such innovations must be sufficiently widely distributed to make plausible the claim that they belong to the entire group; at a minimum, they must occur in languages of the eastern and western groups, and not only in contiguous languages.

5.21. Evidence from verb morphology

5.21A. The gerundive in ***-(i)tavya***

The OIA future passive participle in **-(i)tavya** was a relatively late development in OIA, as far as the texts can tell us. No instances occur in the Rigveda, and only two in the Atharvaveda, though it became fairly common in the Epics (Whitney 1950: 346). The usual meaning in OIA appears to be necessity, probability, and the like, as indicated by such examples as **mayāvasyam gan-tavya-m** 'I must needs go'. Occasionally, however, the form is used with a purely future sense, as in **mayāpi sukhena gant-avya-m** 'I too shall go with ease' (MacDonnel 1927: 202).

OIA **(i)tavya** produced MIA **(i)avva**, which turns up in Gujarati and Marathi as **-āv-**, and in the eastern languages as **-ib/āb-**. The Marathi words with this suffix are variable verbal adjectives in **-āv-C** (C = concord suffix), which Bloch calls the "participe d'obligation" (1919: 256). Its use in Old Marathi and Old Bangla is quite similar to the OIA examples given earlier; for example:

OM	tuvāṃ vadh-**āv-**ā (Instr V-āv-C) 'thou must kill' (Master 1964: 138);
OB	tumhe ho-**ib-**a (= Skt. yuṣmābhir bhavitavyam) 'you will be' (ODBL 967)

The modern languages offer the following forms:[24]

Sindhi	chaḍ-**ib-**o 'is given up'
Gujarati	mahre a cɔpḍi wāNc-**w-**i che 'I want to read this book' (Cardona 1965: 134) (lit. "for-me this book desired-to-read is")
Khandeshi	majā kar-**av-**ī 'rejoicing should be done' (LSI 9.3.214)
Marathi	tyā-ne kām kar-**āv-**e 'He should do the work' (lit. "him-by work to-be-done")

*5.21B. Nominal forms based on the -****(i)tavya*** *gerundive*

From this form is derived a verbal noun which functions as an infinitive of purpose. Examples (from LSI):

Gujarati	kahev-**āv-**ā yogya 'worthy to be called ' (LSI 9.2.367)[25]
Marathi	cārāylā (← cār-**āv-**ayā-lā) pāṭhav(i)le 'sent to graze')[26]
Halbi (Bhandara)	tsār-**āv-**ā-le dhāḍlan 'sent to graze' (LSI 7.363)
Oriya	carā-**ib-**ā-lāĩ paṭhāilā 'sent to graze' (LSI 5.2.386)
Avadhi	hãs-**ib-**e joga 'fit to be laughed at' (Saksena 1971: 283)
Maithili	car-**aba-**lai paṭhaulakai 'sent to graze' (LSI 5.2.61)
Northern Bangla	caṛ-**ābā** pāṭhāy dile 'sent to graze' (LSI 5.1.123)[27]
Assamese	sor-**ābo-**lāi pothāi dile 'sent to graze' (LSI 5.1.406)

This form is well-developed in all these languages, and there are some particularly close parallels in usage between the western and eastern languages, as indicated by the following Marathi and Bangla examples (from LSI) (V = gerundive, Vb = finite verb, Obl = oblique marker, PP = postposition, Aux = auxiliary verb):[28]

Vb-V-Obl-Gen + Aux:	M	kar-**āv-**ayā-ce āhe 'is to be done'
	B	kor-**b-**ā-r āche
Vb-V-Obl-Gen + N-Loc:	M	jā-**v-**ayā-cyā veḷ-i 'at the time of going'
	B	jā-**b-**ā-r somoy-i
Vb-V-Obl-PP + Vb:	M	sāng-**āv-**ayā-sāṭhi ālā 'came to tell'
	B	bol-**b-**ā-r jonne ešeci
Vb-V-Obl-(PP) + lāg-:	M	kar-**āv-**ayā-lā lāglā 'began to do'
	NB	kɔr-**b-**ā lāgil

Western Hindi–Urdu, Panjabi, and Pahari have, for the most part, a suffix in **-n**/**ṇ-** (from OIA **-anīya**) in these uses:[29]

W. Hindi–Urdu	carā-**ne** bhejā 'sent to graze' (LSI 9.1.137)
Panjabi	cār-**aṇ-**lāi ghalliā 'sent to graze' (LSI 9.1.648)
Nepali	carāu-**nu-**lāi paṭhāyo 'sent to graze' (LSI 9.4.58)
Lahnda	carāvi-**ṇ** ghalliyā 'sent to graze' (LSI 8.1.272)

Several other inner group languages/dialects, however, have forms which appear to be derived from the **-tavya** gerundive, such as:

Bandeli	carā-**b-**e ke lāne 'for grazing' (LSI 9.1.416)
Bangaru	carā-**v-**āṇ 'for grazing' (LSI 9.1.259)
Rajasthani	carā-**v-**aṇ-nai 'for grazing' (LSI 9.2.67)

Thus the southern and southwestern varieties of the inner group, which are contiguous with languages of the outer group, show the **-tavya** gerundive forms, while their forms lack the **-l-** suffix. This situation is discussed further in **5.4**.

*5.21C. Future forms derived from the -**(i)tavya** gerundive*

The same OIA form is presumed to be the origin of the characteristic future suffix in **-(i)b/ab-** of the languages of the eastern group (ODBL 965ff.), as in:[30]

Bhojpuri	dekh-**ab-**i 'I will see' (Tiwari 1960: 174)
Early Maithili	kah-**eb-**a, kah-**ib-**o 'I shall say' (Jha 1958: 492)
Bangla	jā-**b-**o 'I am going, will go, should go'
Oriya	**j-ib-ɔ** 'you will go'

It is probable that there exists a relationship between these eastern futures and such western forms as the following:

Surati (southern Gujarati)	ja-**va**-n-o chu 'I will go' (P. J. Mistry, p.c.; see also ja-**w**-a 'I will go', ke-**w**-an 'I will say' in LSI IX.2: 385)
Konkani	kha-**v**-co-na 'won't eat' (Ghatage 1968a: 62)

The **-(i)tavya** gerundive is clearly the most plausible source for these forms, both on phonological and semantic grounds. The **v-** future of Surati, according to LSI, coexists with an **s-** future, though P. J. Mistry (p.c.) states that the **s-**future is no longer used in southern Gujarati. There is no evidence that such forms were ever used more widely in the western languages. Marathi–Konkani has no sign of a **v-**future apart from the Konkani negative forms cited above.[31]

Perhaps the best way to describe this situation is to say that the reflexes of the OIA **-(i)tavya** gerundive had the potential to be used as futures, and that this was realized in a number of outer languages, whereas full-fledged future paradigms developed only in the east. Thus it seems likely that the development from gerundive to future was an innovation which began in the east (or in that part of the original outer speech community which corresponds to the modern eastern languages) and spread partly into the west.

5.22. Phonological evidence

There is also evidence of phonological changes shared exclusively by the outer languages, beyond those features which were proposed by Grierson and refuted by Chatterji (see **5.1**). The following may be mentioned (**5.22A–D**).

5.22A. The vowel ṛ

This vowel merged with other vowels (most commonly **i** or **a**, less frequently **u**) in early OIA. Variant forms in Vedic suggest that the merger process had already begun at the time of the composition of the Vedas (Bloch 1965: 35). In Pali, both **i** and **a** are found, as well as **u** (usually in the vicinity of a labial consonant): for example **iṇa** 'debt' (← OIA **rṇa**), **accha** 'bear' (← **ṛksa-**), **pucchati** 'asks' (← **pṛcchati**). In the inscriptions of Ashoka, the treatment of **ṛ** in the southwest (Girnar, in present-day Saurashtra) is most commonly **a** (Mehendale 1948: 2–3; Bloch 1950:49) The inscriptions of the northwest (Shahbazgarhi and Mansehra, now in the Northwest Frontier Province of Pakistan) most commonly show **i** (Mehendale 1948: 2). Kalsi, near modern Delhi, vacillates between **a** and **i**, but agrees with Shahbazgarhi more often than not. The eastern inscriptions, Dhauli and Jaugada (present-day Orissa), most commonly have **a**, though **u** sometimes appears in the vicinity of labials. When **i** appears in this region, the forms are generally identical to those found in Kalsi and in the northwest, and thus are best

regarded as resulting from lexical diffusion from the midland. Thus the east and southwest agree on **a** as the basic reflex of **ṛ**, as opposed to the **i** of the northwest. (See further discussion in **6.12B**.)

The modern languages essentially confirm this picture. As the following examples show, the usual reflex of **ṛ** in the northwest (Panjabi, Sindhi, Hindi–Urdu) is **i**, whereas it is usually **a** in Marathi–Konkani and Gujarati. Bangla and the other eastern languages show both **i** and **a**; most cases of **i** (e.g. B **piṭ(h)** 'back' ← OIA **pṛṣṭi/pṛṣṭa**, cf. H–U **pīṭh**, M **pāṭh**) can be treated as lexical diffusion from the Hindi–Urdu region.

OIA	*NW(P/S/H–U)*	*SW(M–Ko/G)*	*Eastern*	(*Reference*)
dṛḍha 'firm'	H diḍh, dhaṛ	M daḍṇe (vbl. derivative)	B daṛa	CDIAL 6508
dhṛṣṭa 'bold'	P ḍhīṭh	M dhāṭ, dhīṭ	B ḍā̃ṭ, dhiṭ	CDIAL 6875
mṛttikā 'earth'	P miṭṭī	M māṭī	B māṭī	CDIAL 10286
*vṛnk- 'bend'	P vingā	M vāk-ḍā	B bā̃kā	CDIAL 12066, Southworth 1958
kṛta 'done'	H kiyā	M kelā ← *kata-lla	MB kaïla ← *kaya-lla	ODBL 355
mṛta 'dead'	S mu-o	M melā ← *mata-lla	B maṛa, MB maïla ← *maya-ll-a	CDIAL10278, ODBL 355
tṛṇa 'grass'	P tiṇ	M taṇ	Maithili tan-ik	CDIAL 5906
kṛṣi 'ploughing'	Shina kiṣī	Ko kaśi	B kasi	CDIAL 3448

*5.22B. The vowels **i** and **u***

OIA and MIA show length contrasts between short **i** and **u**, and long **ī** and **ū**. In inner languages like Hindi–Urdu, Panjabi, and Lahnda it is possible to find pairs of words which reflect this original contrast, such as H–U **nīlā** 'blue, dark' (OIA **nīla**) : **milā** 'met, obtained' (OIA **mil-** 'meet, encounter'), **sasur** 'father-in-law' (OIA **śvaśura-)** : **kapūr** 'camphor' (OIA **karpūra**). In the outer languages the contrasts **i/ī, u/ū** do not exist. In Marathi, where the distinction appears in writing, the occurrence of the long and short variants is positionally determined: the vowels are phonetically longer in the final syllable of a word (including monosyllables) than elsewhere, unless followed by a consonant cluster: thus **/i/** and **/u/** are phonetically long in **pi** [pi:] 'drink', **dhu** [dhu:] 'wash', **tin** [ti:n] 'three', **sun** [su:n] 'daughter-in-law', and short in **piṇe** [piṇe:] 'to drink', **dhuṇe** [dhuṇe:] 'to wash', **bhint** [bhintə] 'wall', **uṇṭ** [uṇṭə] 'camel'. Contrasts can only be found in the highly educated formal style in which Sanskritic words are reproduced with a spelling pronunciation reflecting the OIA contrasts, for example **pitā** [pita:] 'father' : **sītā** [si:ta:] 'Sita' (name of a classical heroine). See Kelkar (1958: 21) for further examples.

Old Marathi, from the earliest texts, shows no length contrast between the long and short vowels of each pair. Though this is not stated as such in the historical grammars, the description of the distribution of long and short vowels (e.g. short in initial syllables, long in final syllables except when followed by consonant clusters, see e.g. Bloch 1919: 57, 60; Tulpule 1960: 19) indicates that the situation in Old Marathi was not essentially different from that described by Kelkar (1958) for modern Marathi. In Gujarati the situation is similar (see Cardona 1965: 20–1). Konkani generally shows no contrast (see p. 1 of Ghatage 1966, 1968a,b, 1970, 1972), though varieties of Konkani spoken in south India may have reintroduced the distinction in some cases.[32]

The situation in Old Bangla, as described in ODBL, is identical for all practical purposes: "OIA. and MIA. **i**, **ī** in initial syllables, followed by one consonant, retained its quality in Bengali. Following the Bengali habit of length, monosyllables have the vowel long, and polysyllables short. The quantity is ignored in writing" (ODBL 323); "**u**, **ū** followed by one consonant remain ū in NB…" (ODBL 325); "OIA. **u**, **ū** before two consonants = MIA. **u** → Bengali **u**, **ū**" [depending on position in the word] (ODBL 326). Similar statements are also made for Maithili (Jha 1958: 73–4), Oriya (Majumdar 1970: 22–5) and Bhojpuri (Tiwari 1960: 25). In Ahom (Assamese), "All vowels have different degrees of length, which is allophonic according to their position of occurrence in a word" (Goswami 1966: 84).

Given that the positional determination of vowel length is practically identical in the eastern and western languages of the outer group, it would be reasonable to describe this as a shared change, which apparently was complete before the earliest NIA texts.

5.22C. Word accent

Turner has proposed that there existed a stress accent on the word level in late OIA and MIA, which affected most of the NIA languages, and which "fell on the penultimate syllable of the word if it were long; if short, then on the antepenult; probably if that were short, on the fourth syllable from the end" (1975 [1916]: 47). In other words, the accent fell on the last long syllable of a word (not counting the penult), and on the initial syllable if there was no long syllable. Whether or not there existed such an accent, and whatever its nature might have been, it is a useful device to account for certain changes in the quantity (and in some cases the quality) of vowels which occurred between OIA and NIA in Hindi and several other languages. For example, the different placement of the accent (indicated by the italicized syllable) in OIA **cakra*vā*kaḥ** and ***gar*bhiṇi(kā)** accounts for the different treatment of the vowels in the Gujarati reflexes **cakvā, gābhṇī**; cf. Hindi–Urdu **cakvā**, **gābhin/gābhan**.[33]

Turner notes that Marathi appears to be an exception to the general rule, as indicated by words like **masaṇ** 'cemetery, burning ghat' (OIA **śmaśāna**), **āvsā** (OIA **āvāsa** 'abode'): Gujarati **masāṇ, avās** show the forms expected under the

above accent rule. In order to account for the Marathi situation, Turner makes a number of assumptions, which end up with the accent on the initial syllable of the Marathi word except for one set of cases: nouns of the type OIA **upasthānam**: M **vaṭhāṇ** 'chamber', OIA **upākhyānam**: M **ukhāṇā** 'riddle', in which according to Turner the Marathi outcome is a result of the location of the original placement of the OIA accent. These would contrast with cases like OIA **upalakṣaṇam**: M **oḷakhne** 'recognize', which according to Turner's rather complex argument (pp. 60–64) were re-accented on the initial syllable at some time during the MIA period.

In fact, cases like M **vaṭhāṇ** and **ukhāṇā** show the predominant accentual development which is found in the inner languages (Hindi–Urdu, Panjabi, "Lahnda," Gujarati), as well as (variably) some of the outer languages like Bhojpuri. Thus a more plausible solution of these cases is to posit a general initial word accent for Marathi, and to explain cases which escape this rule as the result of lexical diffusion from areas in which the rule of penultimate accent holds sway. In fact, the large majority of the exceptions can be so explained. Turner himself concedes that Marathi has borrowed words from "closely connected languages of the penultimate stress type…on its northern and eastern boundaries", and also gives examples of possible influence directly from "the literary and religious language Sanskrit" (1975: 68). It is a question of which words should be regarded as directly inherited, and which should be treated as diffused from, or influenced by, other languages. Turner would consider the long first vowel in M **vākhāṇ** 'praise' to indicate Sanskrit influence (OIA **vyākhyānam** 'comment, narration'), whereas under the initial accent rule such cases can be considered regular.[34]

In Bangla, precisely the same interplay of the initial and penultimate accent rules is found. Chatterji notes:

> …during the formative period of Bengali, there were two systems of accent which were current in the language, – (1) the pan-Indian system, which by preference placed the stress on a long penultimate, and (2) the peculiarly Bengali system, which sought to bring all stress to the head of the word.
>
> (ODBL 280)

Chatterji suggests that there was considerable alternation between different forms during the period of Old Bangla, but that "…the initial stress had the victory ultimately, and by the end of the Early Middle Bengali period [*c.*1500 CE, p.132], it is very likely that it was active in West Central Bengali and in most Bengali dialects, thus giving to Modern Bengali words their typical forms" (ODBL 282). Following are examples of Bangla and Marathi words which agree with each other in opposition to words of H–U which follow the penultimate rule (examples

from CDIAL):[35]

CDIAL No.	*OIA*	*Marathi*	*Bangla*	*Hindi–Urdu*
8456	prakṣālayati 'washes away'	pākhaḷṇe/ pakhāḷṇe	pākhalāno	pakhālnā
3103	kāleyaka- 'liver'	kāḷij	OB kālejā/ kalejā	kalejā
11383	vardhāpayati 'increases'	vāḍhaviṇe	bāṛāno	baṛhānā
12961	sambhārayati 'brings together'	sāmbhāḷṇe/ sambhāḷṇe	sāmālā	sambhālnā
13021	samvārayati 'puts in order'	sāvarṇe	sāmlānā	sãvārnā
11383	vardhāpayati 'increases'	vāḍhav(i)ṇe	bāṛāno	baṛhānā
2897	*karmāpayati 'works, earns'	kāmav(i)ṇe	kāmāno	kamānā
2877	karpāsa- 'cotton'	kāpus[36]	kāpās	kapās
4054	gardabha-[37] 'ass'	gāḍhav	gādhā	gad(a)hā
5776	*tāmbolika- 'betel seller'	tāmboḷi	tāmli	tam(b)oli
7966	palyāṇa- 'saddle'	pāḷāṇ/paḷāṇ	pālān	palān

Sindhi appears to go along with the inner languages in this matter. Gujarati, which Turner used as his main example of a language manifesting the penultimate accent rule, also shows a few examples of the initial rule, as well as cases of alternation between the two rules:

39 akṣavāṭa- 'wrestling ground'	G ākhāṛɔ/akhāṛɔ (H akhāṛā)
2226 upalakṣayati 'beholds'	G oḷakhvũ (M oḷakhṇe)
10082 mārjāra- 'cat'	G mā̃jar/mãjār (H manjār, M mā(n)jar)
8607(4) prasthāpayati 'sends forth'	G pāṭhavvũ/paṭhāvvũ (H paṭhānā, M pāṭhaviṇe)

It would be reasonable, then, to conclude that the initial accent rule was an innovation of the outer languages, and that while this rule survived in Marathi and Bangla, its effects were largely overwhelmed in Gujarati and Sindhi by lexical diffusion from the neighboring inner languages.[38]

*5.22D. Change **l** → **n***

An original OIA **l** is replaced by **n** in several slightly different phonological environments, primarily among the outer languages. While the conditioning factors vary, it seems likely that these are related changes – which may, incidentally, point to the influence of pre-Aryan languages. Though these phenomena deserve a more detailed investigation than is possible here, the following brief descriptions will indicate their implications for the inner–outer hypothesis.

(1) **l → n /**[+nasal] ___ (i.e. **l** is replaced by **n** after a nasal vowel or consonant) in the standard Marathi future paradigm, which was formed by the addition of an **-l-** suffix to the old present tense:

mi kari-n (← *karĩ-l) 'I will do'
tu karśi-l 'thou shalt do'
to/ti kari-l 'he/she will do'
āmhi karu (pre-NM karũ ← OM karu-n ← *karũ-l)
tumhi karā-l 'you will do'
te/ti/tyā karti-l 'they will do'

The same rule accounts for the second **-n-** in the standard Marathi pronominal form **tyā-n-nā** 'to/for them' (cf. **tyā-lā** 'to/for him', **tyā-cā** 'his', **tyā-n-cā** 'their(s)') which is derived from **tyā-n-la**, a form still found in non-standard Marathi and some forms of Konkani (see Katre 1966: 148; Southworth 1976b: 305). In Macchimari Koli, a form of Marathi (LSI 7.79), the verb **lāg-** 'begin' appears as **nāg-** following a nasal vowel in expressions such as **karũ nāglī** 'began to do'.[39]

(2) **l** → **n** /___ [+nasal] (i.e. **l** is replaced by **n** *before* a nasal) in some Bangla dialects, for example, in Rajbangshi **āchi-n-u** 'I was/we were' (← ***āchilũ**, cf. **āchi-l-i** 'you were', **āchi-l-o** 'he/she/it was', LSI 5.1.65); the same rule accounts for the **-n-** of the first person perfective suffix in some West Bengal districts: for example, **kon-n-u** 'I did' ← ***kor-(i)n-u** ← ***kar-il-um/ũ** (ODBL 545).

(3) The past/perfective suffix (originally **-l-**, see **5.11**) shows alternation between **-n-** and **-l-** in the following cases:[40]

   (a) Khandeshi has **-n-** as verbal perfective suffix (**bol-n-u** 'I spoke'), and **-el-** as participial suffix (**khā-el poi** 'the eaten chapati', see **5.11**).
   (b) Katkari[41] has a past indefinite suffix **-n-** (**biś-n-a** 'I sat'), a Plup. Act. (sg.) suffix **-nel** (**paḍ-nel** 'had fallen), a Plup. Act. (pl.) suffix **-l-** (**pa-l-** 'had fallen'), and a Plup. Pass. suffix **-(e)l-** (**kāp-el** 'was eaten', **kāp-l-** 'was/were eaten' (Kulkarni 1969: 360ff.; see also Southworth 1976b).
   (c) Maithili also has a morphologically conditioned alternation between **-l-** and **-n(e)-** in the perfect forms, for example, Pres. Perf. (intr.) **calā-l āchi** 'has walked'; Pres. Perf. (tr.) **paḍha-ne āchi** 'he has read' (Jha 1958).
   (e) An **-n-** alternant of the past/perfective suffix also occurs in a number of non-standard forms of Bangla, for example, Kharia-Thar **di-n-u** 'he sent', **ni-n-u** 'he took' (LSI 5.1.92).

(4) OIA word-initial **l-** appears frequently as **n-**, particularly in the eastern group. Chatterji (ODBL 545) states that this was a tendency in "all Magadhan [i.e. eastern] speeches". For Oriya, the CDIAL lists alternates in **n-** for about half

of the words with initial OIA **l-**.[42] Even today uneducated East Bengalis often convert initial **l-** to **n-** (author's observation).

The change **l** → **n** at the beginning of a word is most frequent when the initial **l-** is followed by a vowel plus a nasal consonant: examples from ODBL include **nāngɔl** 'plough' (OIA **lāngala-**) and **nun** 'salt' (OIA **lavana-**). An examination of all the items in CDIAL beginning with **l-V-[+nasal]** yields the following percentages for the change of initial **l-** to **n-**:[43]

Language	*l → l*	*l → n*	*l → l/n*	*%*	*N*
Maithili	0	9	1	95	10
Bangla	7	5	5	44	17
Assamese	6	3	1	35	10
Avadhi[44]	6	3	0	33	9
Nepali	12	5	0	29	17
Bhojpuri	3	1	0	25	4
Oriya	13	2	7	25	22
Hindi	20	3	2	16	25
"Bihari"	9	1	1	14	11
Marathi	22	3	1	13	26
Konkani	3	0	1	13	4
Kashmiri	13	1(?)	0	08?	14
Panjabi	19	1	1	07	21
W. Pahari	14	1	0	07	15
Kumaoni	11	0	2	07	13
Sindhi	17	1	0	05	18
"Lahnda"	17	0	1	03	18
Marwari[45]	1	0	0	00	1
Gujarati	19	0	0	00	19

The languages are listed in descending order, according to the percentage of items showing change. Clearly, this change can be considered to have originated in the eastern region, most probably in the Maithili area (the region of the ancient Magadhan kingdoms). The change appears to radiate from there into Nepali (close to Maithili, with considerable bilingualism in the border area), its neighbors Bhojpuri and "Bihari," the eastern languages (Bangla, Assamese, Oriya) and eastern Hindi – and even to some extent western Hindi,[46] beyond which it seems to peter out. Interestingly, Marathi–Konkani shows almost the same percentage as "Bihari," and though the number of cases here is small, an examination of the individual Marathi words suggests that if there was diffusion of this change between east and west it must have occurred at a time when there was a close link between the predecessors of Marathi and Oriya. For example, in the words for 'plough' (CDIAL 11006: OIA **lāngala-**) a final **-r** appears only in Marathi and Oriya **nāngar**, whereas all the other languages have final **-l** (e.g. Bangla **lāŋal/nāŋal**).[47] In the words for 'tail' (CDIAL 11009(2): OIA **lāngula-**) Marathi has only **n-**forms: **nāng(o/ā)ḍā**, **nāngā** 'scorpion's tail', which have their closest semantic link with Oriya **lānguḍa/nānguḍa**

'tail', **nuḍa** 'bee/scorpion sting', rather than to words such as Hindi **lagūl, lagūr** 'tail', **langūr** 'langur'. The non-appearance of these **n-**forms in Gujarati and Sindhi, which are considered to have been exposed to greater influence of the inner languages, is possibly a result of the stigmatization of the **n-** variants as rustic or uncouth by speakers of inner group languages.[48]

In summary, it is unlikely that the various changes of **l** → **n** listed here, which are similar though not identical in their phonetic nature, are independent of each other. Put together, they add to the evidence of linguistic links between east and west, and also suggest probable contact between the speakers of the outer Indo-Aryan languages and the pre-Aryan groups whose descendants still survive as "tribal" groups (such as the Katkari in the west, and the Santals in the east). Further investigation of these phenomena is needed.[49]

*5.22E. Non-initial post-consonantal **h** → **∅***[50]

Bloch notes that the loss of post-consonantal **h** is found in Gujarati and Marathi,[51] and "is no doubt more general than the orthography…leads us to suppose" (1965: 60).[52] He states further that the loss of post-consonantal **h** is "more or less advanced" in Kafiri, Asiatic Romany, and certain dialects of Bengal and Sind, that it is "usual" in Kashmiri and Shina; in Sinhala the change appears to have gone to completion (Bloch 1965: 62; cf. Geiger 1938: 40). Chatterji (ODBL 441ff.) notes that non-initial **h**, whether post-consonantal or post-vocalic, began to disappear from the early Middle Bengali period. Original non-initial post-consonantal **h** has disappeared in modern spoken Bangla, though it sometimes appears in writing.

5.23. Lexical evidence

There is also evidence of lexical items shared exclusively between the eastern and western parts of the outer group. The following items, which include some grammatical affixes, appear in languages of both branches, and are absent or poorly represented in the inner languages. Though some of these items deserve a fuller investigation, the following list gives a notion of the extent of lexical linkage between the eastern and western branches:

(1) Old Marathi **-te**, **tẽ** 'to, for' (dative affix, see Bloch 1919: 200–1; Tulpule 1964: 56; Master 1960: 56) : Oriya **-te** (used only with pronouns **mo** 'I' and **to** 'you', Pattanayak and Das 1972: 55); said to be from OIA **arthena** 'for the sake of' (**artha-** 'purpose').
(2) Marathi **-t** 'in': **hātā-t** 'in (the) hand', **kholi-t** 'in a room'; cf. North Bangla **bhūya-t** 'in a field' (LSI 5.1.327), **bhāga-t** 'in one's share' (5.1.328); probably from OIA **antar** 'inside', CDIAL 356).
(3) Marathi **-tun** 'from inside' (← **-t** [v. prec.] + **ho-** 'be' + **-un** [gerund suffix]): **gharā-tun** 'from inside the house'; Chakma **-tūn** (LSI 5.1.329).
(4) Marathi **ithun** 'from here' (**ithe** 'here'), Chakma **ittūn** (ibid.).

(5) Marathi **kuṭhe** 'where?', Oriya **kouṭhi**, Bangla **kothe**.
(6) Marathi **āi** 'mother', Ahom **āi** (and similar forms in Dardic languages: Brokpa **āi**, Poguli **yei**, Kohistani **yāi**). Probably a Dravidian or pre-Dravidian word, see DED 364.
(7) Marathi **nigaḍ**, **nigḍi** 'Vitex negundo': Oriya **niguṇḍi**, Gujarati **nagaṛ** (CDIAL 7308, see also **3.22A**(4), **9.24C**).

5.24. Glottochronology

Following are the numbers of items from the 200-word Swadesh list (see Gudschinsky 1956) which are exclusively shared between each pair of groups of NIA (western, central/inner, and eastern):

Central and Eastern only	15
Western and Central only	18
Western and Eastern only	12

Since the number of items shared between the eastern and western groups is only slightly (20 percent) lower than the number shared by central and east, which are contiguous and have been so for at least two millenia, these figures support the notion that an earlier link probably existed between west and east.[53]

5.3. Conclusions

Figure 5.4 shows, in the form of a schematic isogloss map, the distribution within the NIA languages of the main linguistic features discussed in this chapter.[54] These features define a core area which embraces Marathi–Konkani and the eastern languages, including "Bihari," while the western distribution of the isoglosses is more fragmented: Sindhi is clearly marginal, while Gujarati is included by three of the five isoglosses (C–E) and partially included by a fourth (B), as well as the merger of **i ī** and **u ū**. The group of languages so defined might be called "South/Eastern Indo-Aryan," as opposed to the remaining languages, which could be characterized as "North-Central Indo-Aryan." Eastern Hindi (Avadhi), apart from having a verbal noun in **-b-** (see **5.21B**), also has traces of the eastern **-b-** in its future paradigms, which mix **-b-** and **-h-** (← OIA **-iṣya-**) forms, for example, early Avadhi **kahi-haũ** 'I shall say', **kah-aba** 'we shall say' (Saksena 1971: 261–2). Thus Eastern Hindi seems to be an intermediate or transitional zone, as Grierson suggested in his earliest formulation of the inner–outer hypothesis (see Masica 1991: 451).

This figure suggests that these two large groupings of NIA languages represent what Masica might (or might not?) call "overlapping genetic zones" (see **5.0**). These isoglosses represent a paradox for Indo-Aryanists: while a change like **ṛ**→ **a/i** is clearly old, it is difficult to accept the **-l-** past as equally old, when it shows up only in NIA with hardly a trace even in the latest MIA. (While it occurs in Sinhala, the origin there may be different, as noted earlier.) Until now, no

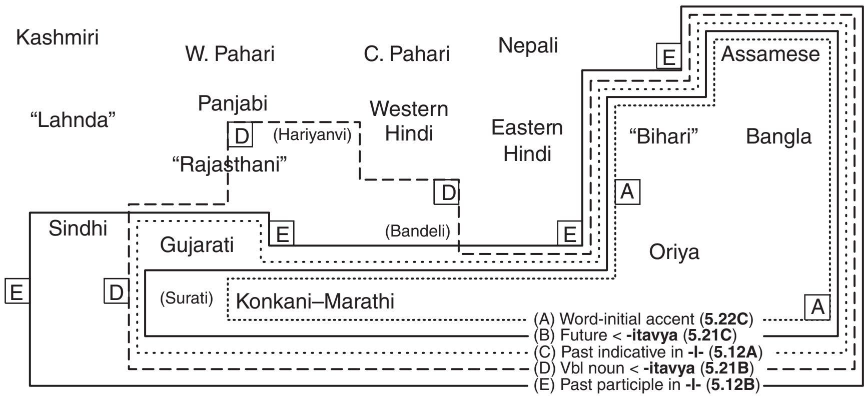


Figure 5.4 Defining isoglosses of "outer" NIA.

Indo-Aryanist since Grierson has supported the notion that the North–Central and South/Eastern language groups represent distinct original subgroupings of Indo-Aryan. The features shown in Figure 5.4 have been dismissed either as independent innovations in the southwestern and eastern groups, or as late developments with no more diagnostic status for subgrouping than any of the other innovations shown in Figure 5.2.

However, the sheer number of innovations between east and southwest, including both phonological and morphological changes, would argue against their being totally independent of each other. In addition, close parallels in morphophonemic detail between south and east were noted both for the **-l-** past (**5.13** end) and for the verbal nouns derived from OIA **-(i)tavya** (**5.21B**), suggesting that these were shared developments. It has already been noted in **5.13** that it is possible to trace the diffusion of some of these changes: for example, the **-l-** past and the word-initial accent appear to have originated in the southwest, while the **-v-** future forms and the change **l** → **n** probably originated in the east. (More accurately, these changes appear to have originated in those areas which correspond to the modern southwestern and eastern languages respectively – wherever those languages might have been located at the time.) The infinitive of purpose in **-v-**, if indeed it originated in the South/Eastern group, has diffused into adjacent North-Central areas. Thus, at least in these cases, we are not dealing with innovations which occurred simultaneously throughout the areas in which they are now found, but rather spread gradually over time.

Furthermore, we cannot assume that the North–Central and South–Eastern regions were ever separate speech communities; it is probable that they were always in contact in some regions, since various other innovations – which we are currently unable to date – have crossed the boundary between these two groups. (See further discussion in **6.4**.) The situation may have been similar to the

relationship between OIA and Iranian (i.e. between Vedic Sanskrit and Old Avestan on the one hand, and between the Yajurveda and Young Avestan on the other) mentioned by Witzel: though OIA and Old Iranian did eventually separate, the parallels between them seem to indicate that, though already differentiated in some features, they underwent a long period of coexistence before separation (Witzel 1989: 237–8).

If independent innovation is eliminated, there would be a need to explain the unusual geographical configuration shared by these changes. What might have been the sociolinguistic or socio-political forces that could have channeled these changes in such a way that they diffused between the southwest and the east, and (with one exception) did not diffuse into the North/Central area? How is it that (apparently) later innovations, such as the **-l-** past, show the same distribution as the change of OIA **r̥** to **a/i** which was attested much earlier? If, during the period of roughly two millennia represented by this series of changes, there was greater communication *within* each of these regions than there was *between* them – as the distribution of these innovations seems to indicate – it is difficult to believe that such large-scale linguistic changes are unrelated to events taking place in the real world. This is as far as the modern linguistic evidence can take us. In order to understand the implications of this evidence for the history and prehistory of the subcontinent, it is necessary to put it into its historical context. This task is undertaken in Chapter 6.

5.4. Summary

This chapter deals with the subgrouping of Indo-Aryan languages. Historical linguists disagree on the division of Indo-Aryan languages into different subgroups, because of a crisscross pattern of linguistic innovations which can justify, or conflict with, almost any attempt to establish areas which exclusively share particular features (Figures 5.1 and 5.2). A division proposed by Sir George Grierson in the early decades of the twentieth century (Figure 5.3), which links the eastern languages with the southern and southwestern languages, has not been accepted by most scholars. The evidence for Grierson's division rests primarily on the modern Indo-Aryan languages, and the features which are shared by the eastern and south/western areas (see Figure 5.4) have been regarded by others as either independent or late innovations. This chapter argues that the number of detailed similarities makes independent innovation unlikely. Whether early or late, the peculiar geographical distribution of these innovations calls for some historical explanation, which is discussed in Chapter 6.

Notes

1 Cardona notes these mergers as characteristic of both the eastern group (Bangla, Assamese, Oriya) and the western–southwestern group (Gujarati, Marathi, Konkani), but does not suggest any connection between them (1974: 448).

2 Strictly speaking, the **s-** future is not a relevant criterion for subgrouping, since it is a retention from OIA (see below and **1.22B**).
3 Lahnda (also sometimes Lahndi) is a name given to western Panjabi, now spoken in Pakistan.
4 See for example Chambers and Trudgill (1998), with references to earlier work.
5 The main exception is isogloss C (**kṣ** > **ch**) in Figure 5.2, which unites Marathi and Gujarati, as Cardona's and Turner's subgroupings do (Figure 5.1).
6 In fact, in the present case we also have the evidence of earlier stages of Indo-Aryan, especially the inscriptions of King Ashoka from the third century BCE (see **6.11**).
7 See Masica 1991: 447–62 for a general discussion of these and other subclassification schemes, with bibliographical references.
8 See Masica 1991: 447–53 for the history of Grierson's views on this question.
9 The languages of the extreme northwest, the so-called Dardic languages, must be excluded from consideration here for several reasons. First, though some of these languages show the past in **-l-**, it is not clear that this past form originated in the same way as in the other languages. Second, the descriptive materials available on most of these languages are inadequate to deal with them in detail. Third, the viability of Dardic as a genuine subgroup of Indo-Aryan is doubtful, since there is no clear evidence (as yet) of shared innovations uniting this group. It seems possible that these languages may represent groups which sought refuge in this mountainous area at various times, and that the similarities among them may result from subsequent convergence. Much more work needs to be done on these languages before firm conclusions can be reached. (The examples in (A) and (B) include data from two northwestern languages, even though these are excluded from the present discussion, as noted here.)
10 Except where indicated otherwise, the linguistic forms cited in this chapter are from the following sources: Assamese: LSI 5.1; early Assamese: ODBL; Bangla: ODBL; Bhojpuri: Shukla 1981a; Gujarati: Cardona 1965; Khandeshi: Chitnis 1964; Konkani: Ghatage 1968a; Magahi: LSI 5.2.123–324; Maithili: Jha 1958; early Maithili: ODBL; Marathi: Kavadi and Southworth 1965, Kelkar 1958; Old Marathi: Master 1964; Oriya: Pattanayak and Das 1972; early Oriya: ODBL; Sindhi: Trumpp 1970.
11 Gujarati has both the **-y-** past (**gayo** in the example here) and the **-l-** past (**gayelo**). Cardona points out that the **-l-** past, which he calls the "perfective verbal adjective" occurs in both attributive position (as in **māhrā bhāie lakhelo kāgaḷ** "my brother written letter" = 'letter written by my brother') and in predicate position (as in **māhrā bhāine ek kāgaḷ lakhyo/lakhelo** 'My brother wrote a letter'). In the latter position, the two forms are interchangeable "for many speakers" (Cardona 1965: 135). However, in a personal communication, Cardona says the following: "…the type [**ga-el-ũ**] is not used in Old Gujarati as a perfective form…this is not the "regular past" formation in modern Gujarati either. To be sure, it can be used in some areas alternatively with the usual form in -y-, but – at least in the standard language – it serves to form an attributive, of the type Hindi [**kiyā huā**]."
12 According to Grierson, the Halbas, who constitute one of the "principal tribes of the Bastar and Kanker States," are probably "an aboriginal tribe" which adopted Hinduism and the Aryan language at some unknown period in the past (LSI 7.330–1). Though Grierson treats Halbi as a dialect of Marathi (under "Broken Dialects of the East," LSI 7.330–91), he describes their speech as a "curious mixture of Oriya, Chattisgarhi, and Marathi" (p. 331). His specimens bear this out: for example, the specimen from Bhandara (pp. 339–40) has clearly Marathi forms like the past **gelā** 'he went' and the gerundive **bharāvā** 'should be filled' (see **5.21**), while Ellichpur (7.369) shows inner-group forms such as **gayo** and **bharno**.
13 The most productive form of the past participle in OIA was **-(i)ta-**, as in **ga-ta** 'gone', **kṛ-ta** 'made', **sthi-ta** 'stood' (from **sthā-** 'stand'), **sthāpi-ta** 'made to stand' (from

sthāpaya 'cause to stand'), **kamp-ita** 'trembled'. In MIA, intervocalic **-t-** → **-y-**, and the sequences **-aya-**, **-iya-**, in this environment as in others, generally ended up as **-e-** or **-i/ī-** in NIA. Thus in Panjabi **geā** 'gone', the **ge-** represents OIA **gata-**, MIA **gaya-**, and the final **-ā** is an agreement suffix which has been added analogically to the inherited form. As noted in **5.12B**, the Marathi equivalent is **gelā** (masc.) **geli** (fem.), in which the **-l-** past ending and the agreement suffixes have been added to a form descended from the original OIA past participle with **-ta-**.

14 See Masica 1991: 180–2 for a discussion of this change.

15 The term "Bhili," like "Halbi" (see note 12), seems to cover a wide variety of different forms which show different relationships to the inner–outer distinction. (This is not surprising, given the location of the Bhil area on the border of 'inner' Rajasthani and 'outer' Gujarati and Marathi.) For example, Grierson cites **goyo** 'went' in District Khandesh (LSI 9/3.103), and in District Nasik **gayā** 'went' alongside of **dinā** 'was given' and **paḍ^{a}n**ā 'fell' (LSI 9/3.146); Kulkarni (1976) cites forms (mainly intransitive) with **-n-** (**uḍ-n-ā** 'flew', **paḍ-n-ā** 'fell', **ā-n-ā** 'came') and both transitive and intransitive forms with **-ø-** (**gɛ̃** 'went', **sāng-ā** 'said', **moḍ-ā** 'broke'). (This **-n-** may be from an older **-l-**; see **5.22D(3)**.)

16 In modern Marathi **ke-l-el-e kām** 'done work', **ge-l-el-ā māṇus** 'gone man'; a suffix **-el-** appears between the past/perfective **-l-** and the concord suffix. It is probable that this **-el-** is an analogical formation based on the relationship existing in languages like Khandeshi and Gujarati between the predicative forms (G **ga-y-o** 'went') and the attributive forms (G **ga-(y)el-o** 'gone'). Something similar probably took place in the pluperfect forms of Konkani: cf. **nidda-lo** 'slept': **nidda-l-il-o** 'had slept'. This would thus seem to be a set of innovations shared by Gujarati, Khandeshi, and Marathi–Konkani.

17 Modern standard Bangla is the exception among varieties of Bangla, and among the eastern languages, in using non**-l-** forms in the attributive position. The B example **gelo bɔchɔr** is cited in ODBL 956 as an exception to the general Bangla rule, and contrasts interestingly with the Gujarati **gai warśe** 'last year', in which the *non* **-l-** form in the attributive position is the exception (Cardona 1965: 135).

18 The possibility of Dravidian origin may also be mentioned. In Malayalam the suffix **-uḷḷa** is used in much the same way as the MIA **-illa/alla/ulla-** suffix, to form attributive adjectives from nominal or verbal forms: **kāramuḷḷa** 'pungent' (**kāram** 'saltiness, pungency'), **pokkamuḷḷa** 'tall' (**pokkam** 'height'), **kuḍikyānuḷḷa** 'potable' (**kuḍikyān** 'for drinking'), **oṭiñña kompu uḷḷa paśu** ("broken horn **uḷḷa** cow") 'cow with a broken horn' (Moag 1994). Both the particular verb form and the usage are very old in Dravidian. The base verb **uḷ** 'be, exist, dwell' (DEDR 697) is probably connected to the noun **uḷ** 'inside, house' (DEDR 698); both can be reconstructed to Proto-Dravidian – as can the construction exemplified by the Malayalam citations, consisting of a relativized verb phrase used in attributive position (Subrahmanyam 1971: 233).

19 In the absence of information about the geographical provenance of these purported midland **-l-** forms, it is pointless to speculate on their origin. On the other hand, it would not be surprising if such forms occurred in border areas as a result of later diffusion, as seems to have happened with the gerundive in **-(i)tavya** (see **5.21A**).

20 As George Cardona has pointed out

> … an affix **-ilá** forming adjectives equivalent to adjectives with **-mat-/-vat-** is known from Sanskrit times, witness Pāṇini 5.2.100:
>
> **-lomādipāmādipicchādibhyaḥ śanelacaḥ**, according to which one has derivatives like **picchila** meaning 'one with a tail'. See also Debrunner, Die Nominalsuffixe 363f…Now, if one has a derivational suffix **-ilá-** that is known to have existed in Indo-Aryan from OIA times, this of course could continue in use at later times.
>
> (George Cardona, personal communication, July 1991)

21 The contracted forms occur from the earliest Marathi. At a certain point these forms were no longer perceived as containing an Aux, and new sequences of Verb + Aux were created, as in NM **gelo āhe** 'I am gone', **gelā āhāt** 'you are gone'. OB has similar forms, for example, **acch-il-esa** 'thou wert', ODBL 947.) See Bloch 1919: 255–57, Master 1964: 122–4.

22 Furthermore, Chatterji points out that although **-l-** pasts with personal terminations are found in OB, "… the old adjectival nature is sufficiently present, making the form take up the **-i (ī)** affix of the feminine when the subject is feminine for the intransitive verb and when the object is feminine for the transitive" (ODBL 947).

23 Note also the form **ās-il-o** 'was' in Brokpa (a northwestern language, LSI 8.2.194).

24 Only western forms are given here, as the eastern reflexes of the **-tavya-** gerundive are either verbal nouns (see below) or future forms.

25 See also Cardona 1965: 133–6, with additional examples such as **huN kām karwā(ne) gayo** 'I went to work'(p. 136, §5.26.4), in which the form **kar-w-ā** is the 'inflected infinitive' (§5.26) and **-ne** is the object marker (p. 143, §6.2.2). Thus the Gujarati form **kar-w-ā-ne** is equivalent morph for morph with Marathi **kar-āv-ayā-lā** [= Verbstem-Gerundive-Oblique-Objmarker] (see the LSI example **cārāylā** in the text here). Note also Gujarati **(kām) karwā lāgyo** 'began to do (work)' (p. 121, §5.20.2a), parallel to the Marathi and Bangla examples given next.

26 **karāvayālā**, **karāvayācā** are contracted to **karāylā**, **karāycā** in modern standard Marathi.

27 While Grierson's "Calcutta standard" Bangla shows a different form (**car-āi-te pāṭhāiya-dilen** 'sent to graze'), the **-b-** form is widespread in rural Bangla: in addition to the northern form (Dinajpur District) cited here, Grierson cites **caṛ-bā-r paṭeya-dil** in the Rajbangshi dialect of Goalpara District (LSI 5.1.169).

28 Chatterji (ODBL 966) believes that the form of the gerundive which occurs in the east is a western form, suggesting that the **-tavya** gerundive may originally have been a western innovation. This is unlikely; see **6.13** for further discussion.

29 This form is presumably derived from the OIA **-anīya** gerundive, which like the **-tavya** gerundive is not found in the RV and is found in only two cases in the AV, but becomes more common in the later language (though not as common as **-tavya**); see Whitney 1950: 346.

30 As compared with the OM and OB forms cited in **5.21(A)**, which conserved the passive character of the suffix, the **-b-** future in the modern eastern languages has become an active form.

31 The Marathi **-āv-ā** conjugation is somewhere between a verbal and a nominal form, in that it takes certain secondary verbal suffixes, such as the 2sg. **-s** in cases like **tu zāve-s** 'thou shouldst go'. Jules Bloch (1919: 253) calls it a "temps…en voie de formation." However, the role of future was taken over in the eleventh century by new future forms based on the old present (Marathi) or the old present participle (Konkani).

32 Ghatage (1967) offers examples of an apparent contrast between **i** and **ī** (p. 3) and between **u** and **ū** (p. 5) in the initial syllables of polysyllabic words in the Konkan of Cochin (in Kerala), but in all cases the short vowel is followed by a consonant cluster. Katre (1966) appears to show a similar situation, though there is no discussion of phonemic contrast.

33 Long vowels shorten when unstressed; VC_1C_1 ends up as long vowel plus C_1 under stress, short V plus C_1 when unstressed. OIA intervocalic [non-retroflex] stops are generally lost, and OIA final vowels are lost except to the extent that they contract with preceding vowels, as in **cakravāka** → MIA **cakkavā(y)a-** → G **cakvā**, ***garbhiṇikā** → MIA ***gabbhiṇi(y)ā-** →) G **gābhṇi**. For loss of final vowels, cf. OIA **mārjāra** 'cat' → Pkt **majjāra** → H **manjār**, G **manjār/mãnjār**, M **mā(n)jar** (T10082), OIA **bhaginī** 'sister' → Pkt **bha(g)iṇī**, **bahiṇī** → H **bahin/bahan**, G **b(ah)ɛn**, M **bahiṇ** (T9349: note that several NIA forms have a final vowel, e.g. Oriya **bhaiṇi**, Konkani **bhaiṇi**; Marathi retains the vowel in non-direct forms such as **bahiṇi+cā** 'sister's').

34 In fact, words like **vākhāṇ**, with two long non-final vowels, create a problem for both Turner's accentual theory and the initial accent rule proposed here. Turner's theory leads him to expect **vakhāṇ** (like **vaṭhāṇ**) whereas the initial accent rule would predict ***vākhaṇ**. Turner suggests that the first ā here may have been influenced by the literary Sanskrit original **vyākhyānam**. If the initial accent rule is accepted, it would be the second **ā** which would require explanation: in this case we might regard the Marathi **vākhāṇ** as a blend of the expected ***vākhaṇ** with the **vakhāṇ** of Gujarati, Hindi, etc.

35 The eastern languages, other than Bangla, have not been studied adequately in this respect. Though the majority of the examples in CDIAL seem to suggest that the antepenultimate stress was the rule in these languages, a number of examples of initial stress can also be found, for example CDIAL 39 (see before) A **ākharā**, B **ākhaṛa**, O **ākharā**; CDIAL 1275. **āmravāṭaka-** 'hogplum', A **āmarā**, B **āmṛā**, O **ām(b)aṛā**.

36 The **u** vowel, seen in the Marathi form here, is also found in Konkani **kāppusa** and in Sinhala **kapu**. (Possible influence of M **kāpur** 'camphor'?)

37 The Hindi **gad(a)hā** (the same form occurs in Maithili, Bhojpuri, and Magahi) implies a precursor ***gardhabhāka-**, in which the long penult would attract the accent in the inner languages, but not in those with initial stress: both the B and M words could also be derived from such a form.

38 A more detailed view of the history of this change might possibly be obtained by a careful analysis of all such forms in the relevant languages, including Panjabi, Hindi–Urdu–Rajasthani, Sindhi, the remaining eastern languages, and even Sinhala.

39 This **-ũ** represents the OIA infinitive ending **-tum** (see Tulpule 1960: 79; Master 1964: 140).

40 While the conditioning factor here appears to be grammatical rather than phonological, it is possible if not probable that these alternations are linked to those discussed under (1).

41 The Katkari language is described in Kulkarni (1969), which characterizes the Katkaris as jungle-dwellers and hunters who inhabit an area in the northwestern ghats. Their language has been classified as a dialect of Marathi, but it is not clear to what extent the language, like that of the Bhils, may incorporate pre-Aryan structural elements. It is also not clear to what extent Katkari has been structurally influenced by standard Marathi, but the probability is that extensive bilingualism has had its usual effects.

42 Chatterji also gives examples of the reverse change **n-** → **l-**, noting that this is "looked upon as a rustic trait" and that "Certain tracts…are noted for the preference among the masses for the [l] sound initially" (ODBL 546).

43 Names of languages in Grierson's outer group are in bold type; those in the intermediate group are in italics. Each case of **l** → **n** is counted as one, and each case of **l** → **l/n** (i.e. **l** and **n** in alternation) is counted as one-half. The percentage is calculated by dividing the total of **l** → **n** and **l** → **l/n** by the total number of cases: thus for Hindi, (3 + 1)/25 = 4/25 = 16/100 or 16 percent.

44 Avadhi = Grierson's "Eastern Hindi."

45 A variety of "Rajasthani."

46 It should be noted, however, that the "Hindi" entries in Turner may include some eastern Hindi items.

47 The "Hindi" **nāgar** listed in CDIAL 11006 is not found in the standard dictionaries of Hindi or Urdu, and thus is probably an eastern or southern form diffused from Oriya or Marathi.

48 In Sinhala, the change **#l** → **n** occurs in three of six cases attested in CDIAL (10905 OIA **langhayati** 'leaps over', Si. **naginavā** 'to rise'; 10917(8) ***laṇḍa-**, Si. **naḍaya** 'dirt; 11006 OIA **lāngala-**, Si. **nagala** 'plough') while the **l** is retained in the remaining three cases: 10951 OIA **lamba-** 'pendent', Si. **lamba**; 10978 OIA **lavaṇa-**, Si. **luṇu** 'salt'; 11154 OIA **loman-**, Si. **lom** 'hair'. Maldivian (an offshoot of Sinhala) also has

several cases of the change. Geiger (1938: 66) states that "the change **l** → **n** belongs to Pre-Sinhalese time" on the basis of its occurrence in Pali.

49 Since some cases of **#n-** are from an original **#r-** (e.g. Bangla **nāch** ← **lāch** ← OIA **rathyā-**, v. CDIAL 10609, ODBL 258), the change **#l-** → **#n-** must have been subsequent to the change **#r-** → **#l-** (see **6.16**).

50 For the purposes of this discussion, "post-consonantal **h**" refers to the aspiration of consonants (transcribed as **bh**, **kh**, etc.); "initial post-consonantal **h**" refers to the aspiration of word-initial consonants, as in Marathi **bhāu** 'brother' **khās** 'special'.

51 Bloch cites G. M. **samaj-** 'understand': H. **samajh-** ← OIA **sambudhyate**; M. **śik-** 'learn': G. **śikh-**: H. **sīkh-** ← OIA **śikṣate**. Following are additional examples showing **h** → **∅** in Marathi and one or more eastern languages. (Gujarati has **h** in some cases, **∅** in others. The loss of non-initial **h** seems to be particularly frequent in words with initial **s**.)

(1) **hāt** 'hand, arm': Ko. **hātu**: A. B. **hāt** (H. G. **hāth**); Pali **hattha-**, OIA **hasta-**
(2) **diḍ** 'one-and-one-half': Ko. **deḍ**: A. **ḍer**: B. **deṛ** (H. **ḍeṛh**) ← OIA **dvyardha-**
(3) **aḍic** 'two-and-one-half': Ko. **aḍḍeca**: A. **ārāi**: B. **āṛāi** (H. **ḍhāī/aḍhāī**: G. **aḍ(h)ī**) ← OIA **ardhatṛtīya-**
(4) **sāḍe** 'plus one-half': Ko. **sāḍe**: B. **sāṛe**: G. **sāṛā** (H. **sāṛhe**) ← OIA **sārdha-**
(5) **sung-** 'smell': B. **sũg-** (G. **sũgh-**, H **sũgh-**) ← OIA ***śṛnkh-**
(6) **soji** 'clean' (of cooked rice): G. **sojũ**: B. **sojā**: A. **xozā** (H. **sojhā**) ← OIA **śodhya-**
(7) **śiḍ(h)i** 'ladder': G. **siḍi**: B. **siṛi** (H. **sīṛhī**) ← OIA ***śrīḍhi-**
(8) **śeṭ** 'respectful term for banker or merchant' (B. H. **seṭh**, G. **śeṭh**,) ← OIA **śreṣṭhin-**)
(9) **śij-** 'cook by boiling': Ko. **śij-**: B. G. O. **sij(h)-** (H. **sījh-**) ← OIA **sidhyati**
(10) **thāmb-** 'stop': B. O. **thām-** (H. **thāmb(h)-/thām(h)-**) ← OIA **stambhate**

52 Bhagwat's grammar of Varhadi shows many examples of loss of **h**, both post-vocalic and post-consonantal: for example, **navtā** was not' (standard Marathi **nahvtā** ← **na hotā**), **buḍi** 'old woman' (Hindi **būḍhī**), **jalā** 'became' (SM **jhālā**), **mane** 'said' (SM **mhaṇe**), **ningun** 'having-departed' (SM **nighun**), **rāylā** 'stayed' (SM **rāhilā**) (Bhagwat 1967: 87).

53 The glottochronological data were collected from educated native speakers of Hindi (Delhi), Marathi (Pune), Gujarati (Ahmedabad), Oriya (Puri), Bangla (Calcutta), and Assamese (Gauhati). The higher number of shared items for the western and central groups may be due to the selection of a western variety of Hindi.

54 The change OIA **ṛ** → **a (5.22A)** has approximately the same distribution as the word-initial accent, shown by isogloss (A) in the figure, while the mergers of **i/ī** and **u/ū** **(5.22B)** have the same distribution as the past indicative in **-l-**, indicated by isogloss **(C)**.

6

HISTORICAL IMPLICATIONS OF THE INNER–OUTER HYPOTHESIS

Contents

6.0. Introduction

Chapter 5 shows the evidence provided by the modern Indo-Aryan languages for dividing modern Indo-Aryan into two sociolinguistic regions, which can be called

the inner (or North–Central) and the outer (or South–Eastern). The present chapter explores the possibilities of reconciling this conclusion with other historical knowledge, namely:

(1) the previous *linguistic* history of Indo-Aryan, that is, OIA and MIA (**6.1–6.2**);
(2) the prehistory and social history of the region (**6.3**).

6.1. Compatibility of the Grierson hypothesis with the earlier linguistic history of Indo-Aryan

The negative reception which the inner–outer hypothesis has received from historical linguists is in large part a result of (1) the belief that the Ashokan inscriptions of the mid-first millennium BCE provide evidence of a different dialect division (between east and west), and (2) the lack of any clear evidence in OIA and MIA which would corroborate the hypothesis – in particular, the lack of evidence for the **-l-** past (see **5.12B**) before about the eighth century CE at the earliest.

In interpreting earlier evidence it is important to note that some scholars, for example Chatterji, have assumed (tacitly or explicitly) that Pali and the Prakrits represent a stage intermediate between the earliest Indo-Aryan and the modern spoken languages. Others take the position that, from the Vedic period onwards, there were varieties of Indo-Aryan which were outside the "high" tradition (e.g. Hock and Pandharipande 1976 on "dialects" in the Vedic period; see **2.82**). If this was true in Vedic times, it would have been even more true during the MIA period when the Indo-Aryan languages were spread over a much larger territory. Thus it is reasonable to assume that along with the attested literary Prakrits there were also "colloquial Prakrits" which never appeared in writing.

It is widely assumed that the earliest OIA, the language of the four Vedas and the associated texts (Brahmanas, Sutras, Aranyakas) was without significant dialectal variation. However, Michael Witzel has shown that, at least in the post-Rigvedic period, significant dialectal differences existed, many of which can be localized with some precision (Witzel 1989). This evidence is discussed in **6.16**, to the extent that it affects the present argument.

6.11. Dialects of the Ashokan inscriptions

On the basis of the distribution of linguistic features in the Ashokan inscriptions, Jules Bloch claimed that in the third century BCE there were in Indo-Aryan "trois grandes régions dialectales: Centre et Est, Nord-Ouest, Ouest" (1950: 85), represented respectively by inscriptions at:

(a) (Center–East) KALSI (near modern Masuri in Uttar Pradesh), DHAULI (near the delta of the Mahanadi south of Bhubanesvar), and JAUGADA (further south on the bank of the Rishikulya);
(b) (Northwest) SHAHBAZGARHI and MANSEHRA (Peshawar and Hazara Districts, respectively, of present-day Pakistan); and

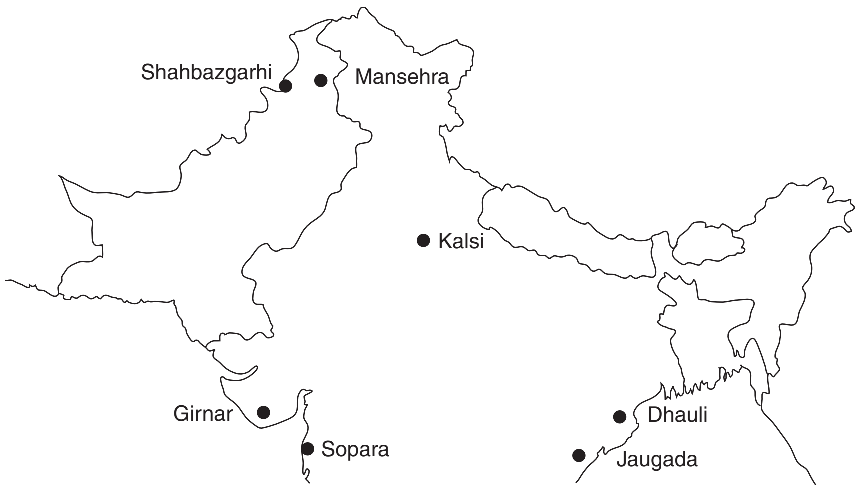


Figure 6.1 Approximate locations of Ashokan inscriptions (after J. Bloch, *Les inscriptions d'Asoka*, Paris: Les Belles Lettres, 1950, following p. 216).

(c) (West) GIRNAR (near Junagarh in Kathiawar) and SOPARA (an ancient seaport on the west coast north of present-day Bombay). (See Figure 6.1.)

There seems to be general agreement that the so-called central/eastern dialect actually represents the speech of the eastern region (Magadha), or rather the chancellery language of Ashoka's court at Pataliputra. According to Bloch, "The midland has, as we know from other sources, a linguistic tradition which is closer to the western regions; nevertheless, Ashoka did not find it necessary to publish his proclamations there in a language different from his own" (1950: 85–6, author's translation). As S. K. Chatterji puts it,

> The language of the Midland does not seem to be represented in the Aśoka inscriptions found within the Midland tract…Here it is practically only a variant of the eastern dialect, which as an official language thus seems in the third century BC to have almost overwhelmed the Midland speech in its own home.
>
> (ODBL 60)

Since some cases of the typical eastern variants also occur in the inscriptions of the western and northwestern regions (see below for examples), it can be assumed that the scribes who drafted these inscriptions probably worked from a Magadhan original. If one discounts these "Magadhisms," Colette Caillat finds that "…the linguistic data available in the Asokan (Major Rock) edicts is basically accurate and trustworthy," in the sense that it truly reflects the local usage of the time (1989b: 430).[1] Therefore we are dealing with three regional variants which will be

designated here as E(ast), W(est), and N(orth). Since the midland dialect is unrepresented, its features will need to be inferred from other evidence (see **6.13A**). Bloch (1950: 43–85) discusses a number of features which differentiate the three major dialect areas, without noting which features represent innovations, and which involve the retention of earlier features. This information, which is crucial to the present discussion, is provided here (**6.12–6.14**).

6.12. East–West innovations

The following four innovations are shared by the eastern and southwestern dialects of the Ashokan inscriptions, but not those of the northwest:

(A) OIA ś ṣ s → E–W s
(B) OIA ṛ → E–W a (N i)
(C) OIA rC_1(h) → E–W C_1C_1(h) (except as in **6.13A**)
(D) OIA C_1r → E–W C_1(C_1)

6.12A. OIA ṣ ś s → E–W s

"…apart from the Northwestern group which retains the three Sanskrit sibilants, Ashoka has a single graphic symbol as in Pali" (Bloch 1950: 48, author's translation). Mehendale (1948: 216) indicates that in the later Prakrit inscriptions (which do not include any northwestern examples), only **s** appears except for sporadic occurrences of **ś** and **ṣ**. In NIA one must assume the merger of the three OIA sibilants for all regions except for the languages of the northwest, a number of which still preserve the three-way distinction.

*6.12B. OIA ṛ → E–W **a (N i)***

OIA **kṛta** 'made, done' → E **kaṭa**, W **kata** (N **kiṭ(r)a**, **kaṭa**). As noted in **5.22A**, this change began during the Rigvedic period. Both the Ashokan inscriptions and the modern NIA languages most frequently show **i** in the northwest and **a** in the southwest.[2] The eastern and central inscriptions also have both **i** and **a**, though in the reflexes of OIA **kṛta** (see forms cited earlier), the most frequent of the words with OIA **ṛ** occurring in the inscriptions, the eastern inscriptions agree with Girnar in showing **a** as opposed to northwestern **i**.[3] This agreement between the east and the (south)west cannot be accounted for unless one assumes that both shared the initial change **ṛ** → **(r)a**; cases of eastern **i** can be regarded as probable diffusions from the northwest via the midland. See **6.42** for the historical relevance of this change.

*6.12C. OIA **rC_1(h)** → E–W **C_1C_1(h)**, except as in 6.13A*

OIA **garbha** 'womb, interior' → W–E **gabbha** (N **grabha**); OIA **darś** 'see, look' → E–W **dass** (N **draś-**); OIA **dharma** 'law' → E–W **dhamma** (N **dhrama**); OIA **karman** 'act' → E–W **kamma** (N **krama**); OIA **sarva** 'all' → W **sarva**/**savva**,[4] E **savva** (N **savra**, Bloch 1950: 54). The retention of **r** (with frequent metathesis)

in the Northwest is reflected in the modern northwestern languages, for example, Tirahi **kram**, Khowar **korum** ← OIA **karman** (Turner, CDIAL s.v.).

6.12D. OIA ***C_1r*** *→ E–W* ***C_1(C_1)***

OIA **priya** 'dear' → E–W **piya** (N **priya**); OIA **agra** 'foremost' → E–W **agga** (N **agra**); OIA **tatra** 'there' → E **tatta**, W **tatra/tatta** (N **tatra**), Bloch 1950: 54. Here again, the modern languages of the northwest retain the **r** in many instances: cf. Waigali **bra**, Tirahi **brā** 'brother' ← OIA **bhrātṛ** (cf. H–U **bhāī**, M **bhāu**), CDIAL s.v.

6.13. Other configurations

6.13A. OIA ***r*** *+ dental stop → N retroflex (+* ***r****), W dental, E retroflex*

Hock (1993: 83ff.) notes that the reflexes of **r** + dental stop are predominantly dental in Girnar, and almost exclusively retroflex in both the northwestern and eastern inscriptions (with some cases of **tr**, **ṭr**, etc. in the northwest). Kalsi, in the midland, shows retroflex and dental outcomes in a ratio of 4:1, from which Hock infers that M and W belong to the same dialect and the cases of retroflex in M represent the influence of E (see **6.11**). "Significantly, this distribution agrees well with the Modern Indo-Aryan outcomes of *r* + dental stop discussed by Turner…" (Hock 1993: 84).

Does this imply that N and E shared the change of dental to retroflex stop before **r**? Such an assumption would imply the further assumption that an earlier connection between N and E was later broken by the intrusion of M. Given that no other innovations show a link between N and E, and considering the differences in detail between the two cases (with N retaining the **r**), it is more economical to assume that these changes were independent (but see **6.14A**).

6.13B. ***kṣ*** *→* ***ch****: a North and West innovation?*

Another case discussed by Bloch which might appear to be a shared innovation between W and N is the change of OIA **kṣ** to **c(c)h**: "Dans l'ensemble on peut poser *ch* occidental pour *kh* oriental" (Bloch 1950: 56). The eastern and central inscriptions consistently show **k(k)h**, while the northwest and southwest have both **ch** and **kh**. However, reflexes of OIA **kṣ** in the Shahbazgarhi inscriptions are written with an additional diacritical mark (which Bloch transcribes with an apostrophe, i.e. **ch'**): for example, E **kham-**, W **cham-**, N **ch'am** 'harm' (Pali **kham-**, OIA **kṣam-**); W **chudda**, E **khudda** 'trivial' (Pali **khudda-**, OIA **kṣudra-**); W **chāti**, E **khaṃti**, N **ch'aṃti** 'patience' (Pali **khanti/ī-**, OIA **kṣānti-**). It is probable however that the N and W changes are independent, since the W **ch** ← OIA **kṣ** coalesces with **ch** ← OIA **ch** (both end up as **s/ś** in Marathi,[5] for example), whereas the northwest retains the distinction. In NIA, this innovation appears in"Dardic" and in the southwest (Gujarati, Marathi, Sinhala), but not in the

intervening languages; Masica refers to these two changes as "quite separate" (1991: 460). The diffusion of the eastern **kh** into both W and N may be again the result of the influence of an original eastern text (see **6.11**) – though the usual representation of OIA **kṣ** in western Hindi is **kh**: OIA **kṣetra** 'field' → H **khet**, etc.[6]

6.13C. OIA $C_1y \rightarrow C_1C_1$

Bloch states (1950: 54) that combinations of $\mathbf{C_1} + \mathbf{y}$ tend to assimilate (to C_1C_1) in the west (i.e. our N and W), while the eastern forms show vowel-insertion: thus OIA **apatya** 'offspring' → W–N **apacca**, E **apatiya**; OIA **śakya** 'possible' → W **sakka**, N **śakko**, E **sak(i)ya**. On the other hand, the cluster **vy** is consistently retained in W: OIA **kartavya** 'to be done' → W **kattavyam**, E **kaṭṭaviye**, N **kaṭavo** (= **kaṭṭavvo**); OIA **divya** 'divine' → W **divya-**, N **div(v)a-**, E **diviya-**. The eastern inscriptions show similar vowel-insertion in various other clusters: OIA **rājñaḥ** → W **rāñño**, N **raño** (= **rāñño**), E **lājine**; OIA **pūrva** 'previous; eastern' → W **puvve**, N **pruve**, E **puluve** (see **6.14B** for **r** → **l**); OIA **svāmin** 'lord' → W **svāmi**, N **spāmi**, E **suvāmi**.

It is probable that these eastern forms with vowel-insertion do not represent the usual spoken forms of the time, because there is no trace of them in NIA: cf. Bhojpuri **bācā** 'boy' (Pali **apacca-**, Pkt. **avacca** ← OIA **apatya-**), Bangla **sā̃c** 'truth' (Pali Pkt. **sacca** ← OIA **satya-**), Bangla **pub** 'the east' (Pali **pubba-**, Pkt. **puvva** ← OIA **pūrva-**). The expected reflexes in Bangla from the forms with vowel-insertion would be ***bāe**, ***sāe**, ***puru** (or possibly ***pulu**). The same must be true of the **vy** in the W forms (at Girnar) such as **kattavyam**, **magavyā**, **divyāni**: since the OIA **vy** appears in the northwestern inscriptions as **v(v)** (presumably representing the actual local pronunciation, inasmuch as it departs from the eastern model), it is probable that **vv** variants also occurred in E and W, and that the **v(i)y** forms seen in the inscriptions represent an elite semi-Sanskritic pronunciation. Furthermore, the modern southwestern languages Marathi and Gujarati show **v** (← **vv**) for OIA **vy**, for example in the reflexes of the gerundive suffix **-tavya** (see examples in **5.21A**), though even today Sanskritized forms like **kartavya** occur in literary Marathi.[7]

6.14. Innovations restricted to the eastern region

The eastern inscriptions show a number of unique innovations, more than the other two regions. While these do not bear directly on the inner–outer hypothesis, they do have indirect relevance to two points: (1) to confirm the notion that the occurrence of E features in the inscriptions of the midland does not reflect the actual midland usage (**6.11**); (2) in terms of chronology, it is important to note that if the eastern region was linguistically differentiated to this extent in the third century BCE, then the period of common outer group changes must lie somewhat further back, say in the early first millennium BCE if not earlier.

(A) OIA ñ ṇ n → E n (**6.14A**)
(B) OIA r l → E l (**6.14B**)

(C) OIA final -aḥ → E -e (**6.14C**)
(D) OIA kṣ → E k(k)h (**6.13B**)

As Bloch points out, the first three changes combine to give the eastern inscriptions a close resemblance to the Magadhi of the classical dramas.

6.14A. Merger of ***ñ ṇ n*** *→* ***E n***

The southwestern and northwestern Ashokan inscriptions show a three-way contrast:[8]

ñ-: W **ñātika**, N **ñatika** 'kinsman' (OIA **jñāti-**)
n-: W **nāma**, N **nama** 'name' (OIA **nāma-**)
-ṇ-: W **prāṇa**, N **praṇa** 'living creature' (OIA **prāṇa** 'breath of life', **prāṇin** 'living creature')
-n-: W **dānam**, N **dane** 'gift, generosity' (OIA **dāna-**)
-ññ-: W **añña**, N **aña** 'other' (OIA **anya**)
-ṇṇ-: W N **tambapaṇṇi** 'Tamraparni' (OIA **tāmraparṇī**)

For all of these, E has only **n**: **nātika**, **nāma**, **pāna**, **dāna**, **anna**, **tambapanni**.

In the literary Prakrits the three nasals generally merge into one, represented by orthographic **n** in Magadhi and **ṇ** elsewhere – though the picture is considerably confused by spelling conventions (Pischel 1981: 165ff.). In the modern languages, the contrast between **n** and **ṇ** is absent in the eastern languages, except Oriya, and in Hindi–Urdu.[9] This situation presumably resulted from the following changes:

(i) **ññ ṇṇ** → **nn**
(ii) Initial **ñ** → **n**[10]
(iii) single intervocalic **n ṇ** → **ṇ**
(iv) **-nn-** → **-n-**, reintroducing the contrast between **n** and **ṇ** in intervocalic position. (This change is part of a general change in which $C_1C_1 \rightarrow C_1$, with "compensatory" lengthening of the previous vowel when stressed, which has affected the eastern, midland, and southwestern NIA languages – see Masica 1991: 459.)
(v) **ṇ n** → **n** prevocalically in a region which now includes Bangla, Assamese, "Bihari," Eastern Hindi, Western Hindi, and non-standard forms of Marathi.

The question remains whether there is any connection between the merger of **ñ ṇ n** → **n** in the Ashokan inscriptions and the merger described in (v). Mehendale's survey of Prakrit inscriptions (1948) seems to indicate that the distinction between **ṇ** and **n** survived at least partially in the central (p. 158) and eastern (p. 188) areas up to the early centuries of the CE. Since the situation in Oriya is essentially the same as that in Marathi and Gujarati (initial **n** and postvocalic **ṇ** ← OIA **ñ ṇ n**, postvocalic **n** ← MIA **ññ ṇṇ nn**), it is likely that changes (i–iv) were shared by

languages of the eastern and western groups. It is unlikely that the merger of **ñ ṇ n** → **n** in Ashoka's time directly reflects change (v), since the Ashokan inscriptions show no sign of changes (i–iv). If the eastern **n** of Ashoka is not purely a spelling convention, then possibly it indicates a merger which took place in some part of the central region of Ashoka's empire, that is in Bihar, and spread only later to the present-day area of western Hindi and eastward to Bangla and Assamese. S. K. Chatterji suggests that the contrast between **n** and **ṇ** probably persisted in Bangla up until the fourteenth century CE, at least in parts of West Bengal (ODBL 523–7).

6.14B. Merger of **r** *and* ***l*** *→ E* ***l***

This change, which produced one of the most striking regional differences in MIA, including the Ashokan inscriptions, requires discussion in a wider historical context. The Rigveda, like early Iranian, reflects a (partial) merger of PIE ***l** and ***r**, with both being represented by **r**, except for a small number of cases where PIE ***l** survives as **l** (mostly in the later books of the RV). In the later Vedas and the literature of the late Vedic and Epic periods, as well as in classical Sanskrit, PIE ***l** survives, both in words which occur with **r** in the Rigveda (e.g. RV **raghu** 'light', AV **laghu**; cf. Lat. **levis**, E. **light**) and in words unattested in RV (e.g. **lunāti** 'cuts', cf. Lith. **liáutis** 'be cut off'). Burrow describes the situation as follows:

> The original division must have been such that the Western dialect turned *l* into *r* in the same way as Iranian (being contiguous to Iranian…), while the more easterly dialect retained the original distinction. [Old Iranian retained only **r**.] It was in this latter area that Classical Sanskrit was elaborated…as a modification of the old sacred language of the Vedic hymns. The latter was always the foundation of the literary language, but since after the earliest period…the centre of its cultivation shifted eastward to Madhyadeśa, in its further development it was subject to the continuous influence of the dialectal forms of this region.
>
> (Burrow 1973a: 84, reprinted from T. Burrow, *The Sanskrit Language*, London: Faber & Faber Ltd, 1973)

There are, on the other hand, a number of cases where OIA **l** appears in place of PIE ***r**, and often in the same distribution as the cases mentioned earlier, that is, with **r** attested earlier and **l** later (e.g. RV **rohita** 'red', AV **lohita**; cf. Lat. **ruber**, G. **rot**, etc.). Burrow claims that

> The number of such examples is too small to justify the assumption of an *l*-dialect to account for them. Such an *l*-dialect does in fact occur later in the Magadhan Prakrit, but it was limited to a small area, and this Prakrit cannot account for forms with *l* out of *r* which occur in the later Vedic literature
>
> (1973a: 85)

While the number of cases of OIA **l** ← PIE ***r** is perhaps small in terms of roots, their representation in OIA is not negligible in terms of specific lexemes. Furthermore, many of these cases resulted in the creation of doublets in an early period of OIA (Epics or late Vedic period), which survived into MIA and NIA. This is true of words with both PIE ***r** and ***l**. This suggests, among other things, that these doublets were not purely literary devices, but represented the actual usage of an earlier period in which a kind of supraregional "standard" or *koiné* form of OIA developed from the conjunction of the northwestern and midland dialects.

Was there an **l-**only dialect? If so, where and when did it exist? The Magadhi Prakrit of the classical dramas has only **l** for OIA **r** and **l**, and the same is true of the eastern and central Ashokan inscriptions. If we add the few cases of **l** from PIE ***r** found in the RV and other early texts (e.g. **klośa** 'a cry', as against the usual **krośati** 'cries', cf. Lith. **kraukti**; two cases of **loman** 'hair' in a late hymn, as against the usual **roman**, cf. ON **rögg** 'long hair', Irish **ruaimne**), we are left with what Jules Bloch considered a real dilemma:

> Should it be supposed that in this as in so many other cases the characteristic change in Middle Indian is very old and contemporary with the earliest documents? Or should we recognize in it traces of a vacillation in Indo-European which has been frequently noticed and no doubt accounts for AV **lumpáti**, O. Slav. **lupiti**, Lat. **rumpo**…? None of the modern languages in fact unite **r** and **l**.
>
> (1965: 76, reprinted from J. Bloch, *Indo-Aryan from the Vedas to Modern Times,* Paris: Librarie d'Amérique et d'Orient Adrien-Maisonneuve–Jean Maisonneuve Succ., 1965)

It is perhaps convenient to assume that the change **r** → **l** of the late Vedic period is historically contiguous with that which shows up so dramatically in Ashoka's eastern inscriptions. I would suggest that, although the solution to Bloch's question about the early occurrence of OIA **l** ← PIE ***r** may not be answerable from the perspective of Indo-Aryan alone, the question regarding the status of an **l-**dialect in MIA times can be meaningfully examined in the light of our present knowledge of sociolinguistic variation in contemporary societies. Bloch himself suggests the relevance of sociolinguistic variation to this problem:

> The retention and the reappearance of numerous words with Indo-European **l** warrant the assumption that they survived in actual speech. Their extreme rarity in the Ṛgveda is a sign of style rather than of dialect; they were evidently in familiar use and their relative rarity even in classical Sanskrit is an indication of the strength of the Brahmin tradition. This accounts for the adjustment of an anecdote in the śat. Brāh. III,2. I 23 made by grammatical tradition: the conquered Asuras deprived of

> articulate speech cry out **helavo helav(o)** or **hailo**, **hail(o)**; Patanjali gives **helayo helay(o)**, which would be a barbarism for **herayaḥ**. This also may partly explain the use made by the classical drama of Māgadhi Prakrit, which it reserves for low-class and ridiculous characters.
>
> (Bloch 1965: 75, reprinted from J. Bloch, *Indo-Aryan from the Vedas to modern times,* Paris: Librarie d'Amérique et d'Orient Adrien-Maisonneuve–Jean Maisonneuve Succ., 1965)[11]

In other words, **l-**words were stigmatized as non-Brahmanical and/or as barbarisms, as early as the time of the Brahmanas (late Vedic period). This in itself does not tell us whether there was an **l-**only dialect at that time, but it makes it clear that the bearers of the Vedic tradition regarded **l-**forms as improper. (See Deshpande 1979b:1–10 for a discussion of linguistic attitudes during the Vedic period.)

As Bloch noted (in the first of the two preceding quotes), the modern Indo-Aryan languages do not support the assumption that there was ever a wholesale merger of **r** and **l** in any region of Indo-Aryan, since the large majority of words undergo no change between OIA and NIA. In fact, some of the words with initial **l-** in Magadhi and Ashokan (such as **lājā** 'king' and **lūpāni** 'spectacles', cf. OIA **rūpa** 'form, beauty') are never found with **l-** at any other time or place in Indo-Aryan. The few cases of word-initial OIA **r** → NIA **l** which can be culled from the comparative materials in CDIAL give the picture shown in the following list:[12]

	CDIAL	*OIA*	*Pa*	*Pk*	*Dr*	*K*	*S*	*LP*	*PN*	*A*	*B*	*O*	*Bi*	*H*	*G*	*M*
1	10543	raktālu (Susr)	—	—	—	—	—	—	—	—	—	~	—	—	—	—
2	10550	rakṣas (RV)	—	r	—	—	—	—	—	—	—	—	—	—	—	l
3	10582	rajju (RV)	r	~	r	r	l	l	r	l	l	—	l	l	—	r
4	10590	raṭati (Kav)	r	r	—	—	r	r	—	—	—	r	—	r	r	~
5	10595	rati (VS)	r	r	—	l	—	—	—	—	—	—	—	—	r	—
6	10600	ratna (RV)	r	~	—	—	—	—	—	—	—	—	—	—	—	—
7	10609	rathyā (Yajn)	r	r	—	—	—	—	*l[6]	—	l/n	r	—	—	—	—
8	10648	raśmi (RV)	r	r	r	—	r	~	r	r	r	r	r	r	r	r
9	10688	rājikā (Susr)	r	r	—	—	r	r	~	l	r	r	~	r	r	r
10	10720	rāśi (RV)	r	r	~	—	r	—	r	r	r	r	r	r	r	r
11	10760	rugna (MBh)	l	~	—	—	—	r?	—	—	—	r	—	—	—	—
12	10816	retra (lex.)	—	—	—	r	r	~	r	r	r	r	r	r	r	—
13	10839	roda (AV)	—	r	—	—	—	r	l	r	l	—	—	—	—	—
14	10539	rakta (SBr)	~	~	r	~	~	~	r	~	~	~	~	~	~	~
	695	alakta (R)														
15	10619	*rapaṇa	l	r	—	—	—	r	—	—	—	—	—	—	—	—
	10936	lapana (Sis)														
16	10620	rapati (RV)	~	~	r	—	l	l	—	—	—	—	—	—	l	—
	10935	lapati (MBH)														
17	10624	rabhate (RV)	l	l	l	—	l	l	l	l	l	l	—	l	l	l
	10948	labhate (AitBr)														
18	10631	rambate (RV)	l	l	—	l	l	l	l	—	—	—	—	—	—	~
	10954	lambate (MBh)														

19	10633 10961	rambhati (MBh) lambhayati (MBh)	l	~	l	—	—	r	—	—	—	—	l	—	—	—
20	10642 10974	*rava lava (Mn)	l	l	—	l	r	~	—	—	—	—	r	r	~	r
21	10645 10986	ravate (Vop) *lavati	l	l	~	—	—	—	l	—	—	—	—	l	—	—
22	10990	laśuna (Gaut)	l	l	—	r	—	l	~	l	r	~	~	~	l	l
23	10668–9	*rahala…	r?	—	—	—	—	—	—	—	—	~	r	—	—	r?
24	10745 11067	*riśyate liśyate (Dhatup)	l	—	l	—	—	—	—	—	—	—	—	r	—	—
25	10799	rūkṣa (SBr) lūkṣa (TS)	~	~	—	l	r	r	r	—	r	~	r	~	l	r
26	10822	*revaḍa	—	—	—	l	r	r	r	—	r	—	—	r	r	r
27	10832 11122	rocate (RV) lokate/locate (Dhatup)	r	~	~	r	—	—	—	—	—	—	—	—	—	—
28	10851 11154	roman (RV) loman (RV)	~	~	—	~	~	~	r	n	r	r	r	r	~	l
29	10854	romaśa (RV) lomaśa (TBr)	r	—	—	—	—	—	—	—	—	—	r	—	—	—
30	10866 11165	rohita (RV) lohita (AV)	l	~	l	—	l	~	~	r	~	~	~	l	l	~

The individual percentages are as follows:[13]

	Pa	*Pk*	*Dr*	*K*	*S*	*LP*	*PN*	*A*	*B*	*O*	*Bi*	*H*	*G*	*MK*
Items 1–13[14]	11	15	13	50	17	25	29	50	29	13	33	17	00	21
Items 14–30	69	65	50	56	55	50	44	70	29	50	28	45	72	45

These figures make it clear that, at least in initial position, there is no language or area which has any monopoly on the change **l** → **r**. The combined scores for each area are as given here:

Language group	*1–13*	*14–30*
Eastern (A, B, O, Bi)	0.28	0.41
Southwestern (S, G, M-K)	0.13	0.57
North Central (Dardic, K, L-P, P-N)	0.24	0.49

All this tells us is that in the first 13 items (with only **r** in OIA) the frequency of **r** → **l** is lower in the southwest than elsewhere, and in the remaining items it is higher in the southwest. Apart from the numerical scores – which may not be significant, given the small numbers involved – if we look at the individual items in

1–13, there are only two (1 and 13) which show **l** in the eastern languages and not elsewhere, and in 4 items (4, 8, 10, 12) the reverse is true. Thus it is difficult to make a case for modern **l** ← **r** in the eastern languages being a residue of a wholesale change which took place before Ashoka's time.[15] Bloch suggests that "the centre of the languages with **l** only must have been the area around Benares and Patna," that is, the central area of Ashoka's empire (1965: 76). If so, why are **l-** forms found as far west as Kalsi, and as far south as Rupnath (in present Madhya Pradesh)? It seems highly likely that this **l** (or the lack of contrast between **r** and **l**) was what sociolinguists would call a *sociolinguistic marker* (Labov 1972: 314–21), that is, a linguistic feature which was manipulated, consciously or unconsciously, as a symbol of social identity (in this case, religious and regional identity). That there was awareness of the **l/r** distinction as a marker of regional difference is shown not only by the variants in the Ashokan inscriptions in different locales, but also by hyperforms in the inscriptions of Girnar and Shahbazgarhi, such as **Turamaya** for eastern **Tulamaya** = Ptolemy (Bloch 1950: 46).[16]

In a chapter entitled "Buddhism and Jainism: sociolinguistic self-defense," Madhav Deshpande states that the Buddha and his followers conducted their discourses in the local Prakrit rather than Sanskrit, not only because it was a more effective vehicle for reaching the masses, but because they *considered it superior* to Sanskrit, in fact more Aryan (**ariya**) than Sanskrit:

> On the higher philosophical plane, Buddha totally rejected hereditary caste rank . . . However, on the lower plane, there is a clear assertion that Kṣatriyas are superior to Brahmins. Thus, from his point of view, far from being an inferior dialect, Buddha must have considered his own dialect superior to that of the Brahmins, as he considered his own Kṣatriya rank superior to theirs . . . only on this interpretation can we explain why the Pālī Buddhist tradition came to view Pālī to be the supreme original language of all beings including gods.
>
> (1979b: 41, reprinted from M. Deshpande, *Sociolinguistic Attitudes in India*, Ann Arbor: Karoma Press, 1979, with permission of the author)[17]

Thus it is likely that in Gautama's own speech, and the speech of his area of origin, the pronunciation of **r** was such that it was interpreted by those from other areas as **l**. Possibly there was no distinction between the two, but that is not a necessary conclusion. The village of his birth appears in Buddhist literature under the name **Lumbini**, said to be derived from the name of a princess and a grove named after her (Monier–Williams s.v.).[18] This village, now called **Rummin-dei** or **Rummin-di** (= OIA **rukmiṇī-devī**), is located in southern Nepal. Gautama spent most of his life teaching in the kingdoms of Magadha (centered on Pataliputra = modern Patna) and Kosala (northwest of Magadha) and in the "tribal lands to the north of the Ganges" (Basham 1954: 256–7). Presumably most of his followers, at least in the early stages, were also from this general area. Thus

the most likely location of the **l-** Prakrit would be Southeastern Nepal in the region near the modern districts of Champaran, Birganj, Chapra, Muzzafarpur, and Darbhanga in Bihar.

But Buddhism also attracted followers from distant regions, and it is likely that those who came from areas where **r** was predominant would have learned to imitate the **l-** dialect, not only out of the universal desire to identify with their neighbors, but also, in the context of early Buddhism, to identify with the new religious, political, and social ideology which rejected the caste hierarchy and the hegemony of the Brahmans. And inevitably, as is the case with immigrants to modern urban areas (Labov 1972: 317ff.), this imitation would have been sometimes carried too far, resulting in some cases of original **r** being converted to **l**. It is possible that the few cases of **l** ← OIA **r** which we find in the eastern area today are the residue of such a situation – though, as noted earlier, equal numbers are found in other regions.[19]

Coming to Ashoka's time, it seems possible that the relentless substitution of **l** for **r** which appears in the eastern inscriptions reflects primarily the conscious use of **l** as a sociolinguistic marker to differentiate the Buddhist order from the Brahmanical establishment, perhaps enhanced by the missionary zeal of Ashoka, a late convert to Buddhism. This might be regarded as a type of positive stereotype, indicating rejection of the Brahmanical norms of speech as well as behavior. The use of **r-** forms (including some hyperforms, as noted earlier) in the inscriptions found in the regions outside of Ashoka's direct control was probably not only a response to the Buddha's enjoinder to preach to people in their own tongue, but also indicates the low opinion held of **l-** forms in those regions (see Deshpande 1979b: 115, note 131).[20]

The continuation of the **l-** forms to the exclusion of **r-** forms in the literary Magadhi Prakrit probably has a different, though related, explanation. In the classical dramas, Magadhi appears in the mouths of clowns and lower-caste people, who are not only looked down on as degraded or ridiculous, but also on occasion are allowed to express anti-establishment views. Such viewpoints can of course be most safely expressed by those in the persona of jesters, who are not to be taken seriously, and who may even represent a historical continuation of the anti-establishment role of Buddhists and members of other heterodox groups.[21]

The time interval between the earliest changes of **r** → **l** (in the late books of the RV and the AV) and the time of the Buddha is at least 500 years, perhaps closer to 1,000. If there is any connection between the earlier and later cases, it would most likely have involved a non-IA language peripheral to the eastern region of OIA and MIA, in which OIA **r** and **l** would both be rendered as something closer phonetically to **l** – peripheral because we must assume an area whose inhabitants were never fully assimilated to the Aryan society, but who interacted regularly enough with Indo-Aryan speakers to have a detectable influence on their speech.

Gautama's birthplace, and the location of his Sakya (OIA **śākya**) tribe, was in the foothills of the Himalayas. Possibly one or more of the indigenous groups in this area favored **l-**forms, and may have been responsible for the **l-**forms which

entered Indo-Aryan from the late Vedic period up until MIA times. Newari, a Tibeto-Burman language of the Himalayas, would be one possible candidate for such a role. Though Newari has an **r/l** distinction, it does not seem to be old: in initial position, **l-** is fairly common, but **r-** appears mainly in foreign words such as **rājā** 'king', **rumāl** 'kerchief' (from Nepali or Hindi–Urdu), and **ryadywa** 'radio' (Sresthacharya 1995; see also Genetti 1988: 164).

*6.14C. OIA **-aḥ (~o)** → E **-e***[22]

Bloch notes that the **-e** is consistent in the inscriptions of Kalsi (central), Dhauli and Jaugada (east); that **-o** is the more frequent ending at Girnar and Sopara (southwest) and at Shahbazgarhi (northwest), though cases of **-e** are "assez nombreuses" at these three sites; and that **-e** is more frequent than **-o** at Mansehra in the northwest (where the inscriptions are "notoirement infecté de 'magadhismes' variés," Bloch 1950: 47). Generally speaking, Girnar (W) and Shahbazgarhi (N) agree with Pali in showing mainly **-o** with an occasional **-e** (1950: 47–8).

Mehendale's survey of Prakrit inscriptions shows the **-e** ending primarily in the east, with occasional occurrences in the central and western regions up through the first century BCE (1948: 238). In the NIA languages, the **-e** ending is the norm for both masculine and neuter nouns in Bangla, Assamese, early Maithili, and early Oriya (ODBL 740), whereas the other modern languages show no trace of it. Thus this **-e** ending was probably an eastern innovation which never diffused beyond the eastern region as far as actual speech is concerned.[23]

6.15. Ashokan dialects (summary)

Bloch's three-way division of the Ashokan dialects (Center–East, Northwest, West) can be resolved into an earlier two-way division between the Northwest and the remaining dialects. As shown in Figure 6.2, all but the Northwest (referred to above as N) share the four innovations discussed in **6.12** (OIA ś ś s → E–W s, OIA ṛ → E–W a (N i), OIA rC_1 → E-W C_1C_1, OIA C_1r → E-W C_1C_1). There is overlapping of innovations in the Midland dialect: the change **ṛ** → **i** is shared by the Midland and the North (**6.12B**), the change of **r** + dental → dental is shared by the Midland and the West (see **6.13A**), and the change of **kṣ** → **kh** is shared by Midland and East (**6.13B**). The change OIA C_1y → C_1C_1 (**6.13C**) appears to be common to all dialects; the change of OIA **kṣ** to **c(c)h**, though seemingly shared by N and W, probably represents two independent changes (**6.13B**). The change of OIA **r** + dental stop to retroflex stop may possibly have been shared by N and E, if we are prepared to admit that this was overlapped by a (later?) change to dental stop shared between W and the Midland dialect, but this could perhaps better be regarded as two independent changes (**6.13A**). Several changes are restricted to the Eastern group (**6.14**), and thus may have taken place later than those mentioned here. The modern NIA languages essentially confirm this picture, with the Midland sharing some innovations (probably through

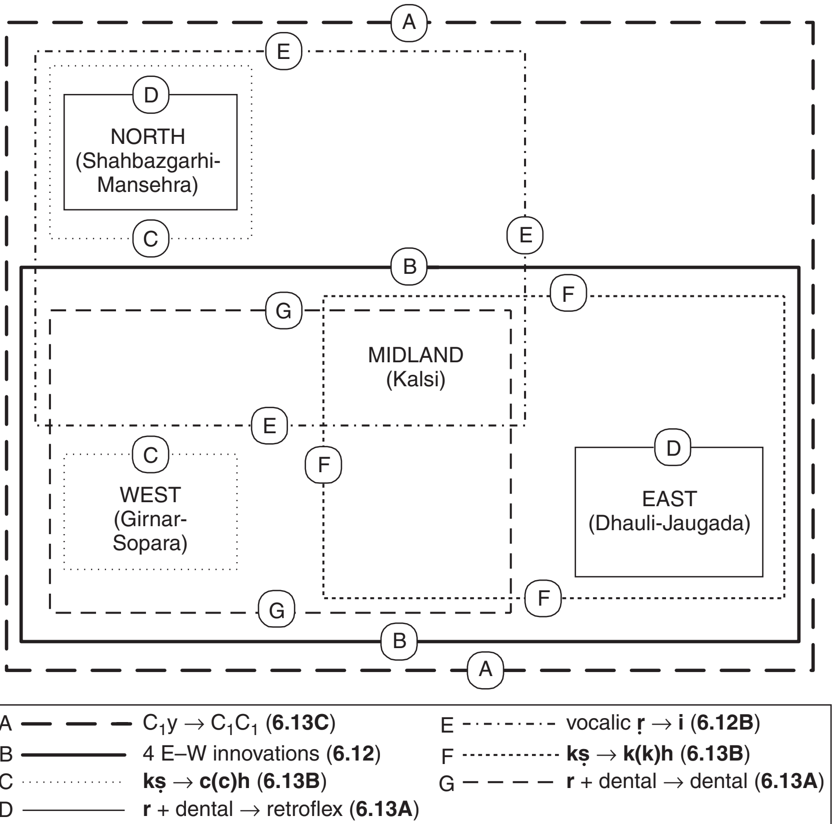


Figure 6.2 Phonological isoglosses in Ashokan inscriptions.

gradual diffusion) with each of the peripheral dialects: OIA **ṛ** → **i** (M and N, see **6.12B**), OIA **kṣ** → **k(k)h** (M and E, **6.13B**), OIA **r** + dental → dental stop (M and W, **6.13A**).

6.16. Vedic dialects

Though he emphasizes that there is much more work to be done, Michael Witzel has mapped the major dialect areas of the post-Rigvedic period, on the basis of a number of innovations which can be traced in texts whose dates and locations can be pinpointed with some precision (Witzel 1989). He concludes that there are three main centers of innovation: (1) the Kuru area (East Panjab–Haryana), (2) the Pañcāla land (western Uttar Pradesh), (3) the East (primarily Videha, i.e. North Bihar); the eastern area was “the late Vedic center of major innovations” (1989: 224–5). In addition, there are two transitional dialects, Kosala, and the

south, which are "…slightly problematic as many developments which took place there also occur in the Eastern area…" (1989: 226). On the assumption that certain texts from these regions are more or less contemporary (an assumption which needs further investigation), Witzel concludes that "…one may posit a late Vedic Eastern Central/ Southern/Eastern dialect grouping which stands opposed to the earlier Kuru-Pañcāla area" (ibid.). Witzel also notes that the Eastern (Prācya) dialect "emerged into prominence only during the late Brāhmaṇa period," and though the Jaiminīya Brāhmaṇa indirectly criticizes a king's sons for speaking like the easterners, "Subsequently, some of the characteristics of the Prācya dialect seem to have been [so] strongly preferred…that they penetrated…westwards until they reached the Panjab…" (Witzel 1989: 226).

On the basis of Witzel's work in locating the various Vedic texts in time and place, it is possible to suggest some correlations between groups of textual schools or cultures and archaeological traits. For example, he points out that the area covered by a group of Yajur Veda and Atharva Veda texts coincides with the area covered by the Painted Gray Ware culture (1989: 247). He also notes that when NBP (Northern Black Polished Ware) spread rapidly throughout northern India (including some areas south of the Vindhyas) around 500 BCE, its area of distribution matches closely with that of the late Vedic Brāhmaṇa texts (ibid. 244).

In the context of the present chapter, the salient information to be drawn from Witzel's study of Vedic dialects is, briefly, that in the late Vedic period, that is not long before the Buddha's birth, two major dialect areas existed, one located in the midland – that is, the eastern Indus and the Ganga–Yamuna doab – and the other combining the east and south, with a transitional area in Kosala, roughly modern Avadh.

6.2. Summary of the linguistic evidence

Evidence for the existence of two distinct sociolinguistic regions, inner (North–Central) and outer (South–Eastern) Indo-Aryan, was presented in Chapter 5. In **5.1** and **5.2** extensive phonological and morphological links between the eastern and western parts of the outer region were shown, and a case was made for the outer languages having shared a long period of common historical development which was (at least partially) distinct from that of the inner languages. The principal point in favor of this conclusion is the wide and continuous distribution of these innovations, which argues against independent creation or recent diffusion. Whatever may be the objections to the linguistic argument, it remains difficult to explain away the phonological and morphological evidence for the links between east and west. The totality of the evidence points to the existence of two sociolinguistic regions, each showing some internal uniformity *vis-à-vis* the other, which however were probably in at least intermittent contact throughout most of their history.

While most of the innovations cannot be dated, the change of OIA **ṛ** to (inner) **i** and (outer) **a** (**5.22A**, **6.12B**) goes back to the Vedic period, suggesting that the

original division could be as old as the early second millennium BCE. Its distribution in the Ashokan inscriptions suggests that it was one of the earliest features differentiating the Northwest from the rest of Indo-Aryan. Apart from lexical diffusion, this distribution has remained constant up to NIA. The **-tavya** gerundive, which makes its first appearance in the Atharvaveda, appears to be a somewhat later feature, which apparently diffused into midland area by the mid-first millennium BCE, as evidenced by its acceptance in Panini's grammar.[24]

The distribution of other features shown in the Ashokan inscriptions (**6.15**) is consistent with an earlier division between Northwest–Central (as represented by Shahbazgarhi–Mansehra and Kalsi) and Southwest–East (Girnar and Dhauli–Jaugada). However, the dating of these innovations is difficult. The most one can say is that those innovations shared between southwest and east would be expected to predate those restricted to the eastern group (**6.14**), all of which occurred before the mid-third century BCE.[25] The same would apply to those changes, such as the loss of length contrast between **i/ī** and **u/ū** shared by the eastern and southwestern languages (**5.22B**), for which evidence comes from NIA.

The **-l-** past adds its own complications to this picture: it is dateable textually to the eighth century CE at the earliest, though if (as assumed in **5.13**) the **l-** participle was the earlier innovation, then the latter might possibly be placed as early as 500 CE. This would still leave a gap of over 1,000 years between this and the earlier change **ṛ** → **a/i**, though the two have very similar distribution in NIA. Thus there are two problems demanding explanation: chronology and distribution. While the **l-** participle *might* be older in the west, there would still be a need to explain how it diffused into the eastern languages without showing up in the North–Central area. Thus it would seem that for at least a millennium, a "channel" remained open which allowed linguistic innovations to pass between the southwest and the east without affecting other areas. This problem will be taken up again in **6.4**.

The evidence of Vedic dialects (**6.16**) does not conflict with, and possibly supports, the inner–outer group hypothesis, in that the major dialect division in the late Vedic period is between a midland dialect and an eastern–southern dialect, with a transitional dialect in the area of Kosala, the modern Avadh – exactly where Grierson placed his intermediate group. The northern Panjab, less active at this time in terms of text production, forms a separate dialect area.

As noted in **5.12**, there is no evidence to support Bloch's claim that **-l-** past forms occur in the modern midland languages. If such evidence should appear, then a different set of questions would arise: if the **-l-** past is widespread throughout NIA, when and where did it originate, and what happened to it in the areas where it no longer exists? Why is there no early evidence of it anywhere? Since its distribution corresponds with that of the early change of OIA **ṛ** → **i/a**, there is a possibility that the **-l-** past existed earlier in speech though not in writing. Clearly this would imply an association with non-elite social dialects, possibly those of Shudras or outcastes – or even groups who spoke OIA as a second language. This could easily explain its non-occurrence in written texts, but for

how long? It is perhaps significant that when it finally does appear in full force around the eleventh century, it is in the context of a somewhat "counter-cultural" tradition in terms of both authorship and content (see e.g. Feldhaus 1983).

This phenomenon raises some interesting methodological questions: first, do we have any reason to assume that it would be impossible for a high-frequency form to remain "underground" in this way for 500–1,000 years? Perhaps we do not yet know enough to answer this question. Second, are there any features of the South Asian situation that suggest that it would be more likely to happen here than elsewhere? Caste barriers might provide a solution: while we do not know exactly how these barriers functioned in the first and second millennia BCE, something relevant can be said on the basis of modern sociolinguistic investigations. Gumperz (1958) describes certain linguistic differences in a north Indian village, and notes that the most marked differences were those between members of touchable and untouchable castes. While some of the features of the untouchables' speech were known to those of higher caste, they were considered stigmatized forms and were ridiculed by higher-caste people. Thus it is unlikely that a high-caste person would knowingly use such forms.

Southworth (1975), reporting research done in South India in the 1970s, noted that members of different castes living in the same village were ignorant of the radically different meanings given to certain words by members of other castes, and only became aware of them in certain modern contexts – for example, when children from different castes interacted in municipal schools. Clearly, the opportunities for such contacts would have been much rarer at an earlier time. The traditional situation in Kerala, as described by K. R. Unni (1959), may give some idea of the historical relations among different castes. Unni notes that there were strict rules regarding distances to be maintained between the various categories of castes, as well as rules governing visual contact and prescribed/proscribed speech. For example, members of low castes, if they did have occasion to speak to higher-caste people, avoided the use of many words connected with personal identity (e.g. pronouns, words for dwellings) and personal habits (e.g. dress, eating). There was in fact very little direct contact between members of low and high castes, since dealings between these groups were generally mediated by others; for example, low-caste laborers on a high-caste family's land would be supervised by an overseer, usually of an intermediate caste. In such a situation it might be possible for a feature of low-caste speech to remain "underground" for long periods of time.

6.21. Further evidence of linguistic links between eastern and southwestern Indo-Aryan

M. H. Klaiman's paper, "Bengali syntax: possible Dravidian influences" (1977), discusses a number of structural developments in Bangla which Klaiman believes to be attributable to Dravidian influence. For example, both Bangla and Tamil (as well as other Dravidian languages) have derived a general complementizer from

a verb meaning 'say': both languages have sentences of the form **Σ** + **Comp** + **VP**, in which the Complementizer element is a participle of a verb meaning 'say' and the VP (verb phrase) contains verbs like 'hear', 'say', 'think': for example, Bangla **Σ** = [āpni ḍhākāy jācchen "you Dhaka-Loc. are-going"] **Comp** = [bole "having-said"] **VP** = [āmi suneche "I have-heard"] 'I have heard that you are going to Dhaka"; Tamil **Σ** = [pāl illai "milk is-not"] **Comp** = [enru "having-said"] **VP** = [avan connān "he said"] 'He said, "There's no milk"'. Klaiman points out that Bangla also has a clause-initial complementizer **je** which occurs in sentences of the type **VP** + **Comp** + **Σ**, which reverse the order of elements from the previous examples: for example, **VP** = [āmi suneche "I have-heard"] **Comp** = [je "that"] **Σ** = [āpni ḍhākāy jācchen "you Dhaka-Loc. are-going"] 'I have heard that you are going to Dhaka'. Since Tamil and other Dravidian languages historically only had sentences of the first type (**Σ** + **Comp** + **VP**), and since some other Indo-Aryan languages – Hindi–Urdu in particular – lack this first type, Klaiman concludes that the creation in Bangla of the clause-final complementizer **bole** was not part of a general early Dravidianization of Indo-Aryan, but rather a specifically Magadhan (eastern IA) innovation of relatively recent date.

One may gain a somewhat different perspective by looking at the other outer IA languages. All of the features cited by Klaiman have parallels in Marathi–Konkani, and several of them occur in other southwestern languages.[26] For example, Marathi has both types of quotative sentences as illustrated earlier: **Σ** = [saraḷ jāun ujvikaḍe vaḷā "straight having-gone right-side turn"] **Comp** = [mhaṇūn "having-said"] **VP** = [tyāni sāngitle "him-by said"] 'He said, "Go straight ahead and turn left"'; **VP** = [tyāni sāngitle] **Comp** = [ki "that"] **Σ** = [saraḷ jāun ujvikaḍe vaḷā] 'He said (that)…'.[27] Another case involves a post-position meaning 'from' derived from a copular verb: Bangla **kolkātā theke** 'from Calcutta' (← **thāk-e** 'having- been/stayed'): Marathi **mumbai-hun** 'from Mumbai' (← **ho-un** 'having-been'): Tamil **ūril-iruntu** 'from town' (**ūr-il** "town-in" **iru-ntu** "having-been").

These cases are reminiscent of the situation which exists with IA nominal endings such as the possessive suffixes: Hindi–Urdu **kā**/**kī**/**ke** (as in **rām kī kitāb** Ram's book'), Panjabi **dā**/**dī**/**de** (e.g. **śyām dī haṭṭī** 'Shyam's shop'), Gujarati **nā**/**ni**/**nū**/etc. (**bhāi-nū ghar** 'brother's house', v. Cardona 1965: 144–5), Sindhi **jo**/**jī** (**ghar jo dhaṇi** 'master of the house', Trumpp 1970: 119), Bangla **-r~-er** (**rām-er** 'Ram's', **ghoṛā-r** 'horse's'),[28] Nepali (invariable) **ko** (e.g. **usko bābu ko ghar** 'his father's house'). In this and other cases, the elements in the different languages show closely similar meanings and functions, though the individual lexical items are clearly not cognate with each other. This looks suspiciously like what creolists have called 'relexicalization', that is, different local replacements of an earlier common lexical item.[29] Basic patterns like **Noun** + **Postposition** and **S** + **Comp** + **VP** (illustrated earlier) may well be proto-outer patterns which developed at an early stage of outer IA (see **6.4**), with later relexification reflecting the influence of local languages.[30]

Thus the similarities noted by Klaiman between Bangla and Tamil appear to be of wider distribution. In fact, some of these features even appear in the inner languages: for example, Hindi–Urdu has **choḍ-ke** ("having left") used as a postposition meaning 'excluding', parallel to Tamil **viṭa** 'leaving', as well as **cāhiye** 'is needed/wanted', closely parallel to Bangla **cāi**, Tamil **vēṇṭum**. Nepali **bhanera** 'having said' has all the uses attributed to Bangla **bole** and Marathi **mhaṇun**.[31] But the inner languages clearly have fewer of these features, and where they have them they appear to be less firmly established. Thus these features may be considered compatible with the hypothesis that the outer languages were exposed to the influence of indigenous South Asian languages from a very early period (see **6.4**), with some features being passed on over time to the inner languages.

On the other hand, we have no reason to conclude that all the features in question are originally Dravidian. They may go back to an even older substratum, since some of them appear in Munda languages also (Masica 1976: 189). What is clear, however, from this account is that a number of these features are shared between Dravidian and outer Indo-Aryan, while the inner IA languages share a much more restricted set.

6.3. Other evidence

6.31. Textual evidence

Aryavarta (OIA **āryāvartta**), the homeland of orthodox Hinduism, is defined in OIA literature as an area corresponding roughly to the western part of modern Uttar Pradesh, bounded on the east by the confluence of the rivers Yamuna/Jamna and Ganga. (See Deshpande 1979b: 47–8 for details.) This is the primary area associated with the late Vedic period, and with the Mahabharata epic (see **2.21B**). Thus the identification of Aryavarta with orthodox Hinduism is reflected in the "mainstream" of OIA Brahmanical literature. Aryavarta was frequently contrasted with the *mleccha-deša*s, areas which were occupied by *mleccha*[32] peoples and not supposed to be entered by devout Hindus.[33] The *ārya–mleccha* distinction involves not only language differences, but also other matters such as kinship, occupation, and ritual status. In general, the lands outside of Aryavarta were considered as *mleccha-deša*, even though many of them – and presumably more of them as time went on – were populated by speakers of Indo-Aryan.[34]

Both the Deccan (the region south of the Vindhyas) and the east (the lower Ganga valley) were counted among the traditional *mleccha-deša*s (Thapar 1973).[35] It was noted earlier that the eastern speech was looked down on by the midland Brahmans, even though some of the features of the eastern dialect later diffused into the midland. Regarding the Deccan, Thapar comments:

> It is curious that in spite of considerably increased communication between the Ganges valley and the peninsula and the spread of Sanskrit and of Aryan culture to the south, there is a persistence in regarding the

> southern regions as *mleccha-deśa*. The Andhras, for example, who had ruled the northern Deccan for four centuries, are described as *mleccha* kings and their lands unfit for the *śrāddha* ceremony. At the same time the Andhra kings were claiming to be the protectors of the varṇa dharma, and the destroyers of the Sakas and Yavanas.
>
> (Thapar 1973: 422, reprinted with permission from 'Image of the barbarian in early India', *Comparative Studies in Society and History* **13**: 408–36, copyright © by Cambridge University Press)

The boundary between Aryavarta and the south is formed by a mountainous belt known as the Vindhyan complex (see Figure 6.3). According to Subbarao,

> The most important belt running right across the country from the West coast to the Delta of Bengal may be described as the *Vindhyan complex* comprising the Satpuras, Vindhyas, Mahadeo hills, Gawaligarh, Maikal range, Hazaribagh range, the Chota-Nagpur, the Singhbhum and Manbhum plateaux. This is also the most populous tribal belt sheltering Bhils, Dangs, Gonds, Santhals, Uraons, Baigas, Gadabas, Marias, and a host of other tribes.
>
> (Subbarao 1958: 14, quoted with permission from B. Subbarao, *The Personality of India*, Baroda: Faculty of Arts, M. S. University of Baroda, 1958)

The boundary between inner and outer Indo-Aryan runs to the south of the Vindhyan complex. The association of hunting-and-gathering groups with this



Figure 6.3 Mountains and hills of Central India with the approximate location of "tribal" languages.

area is clearly quite old, and although speakers of Indo-Aryan and Dravidian have lived alongside "tribal" peoples for many centuries, these hills and mountains surely constituted a much more formidable barrier in early times than they do now. Thus in Ashoka's time, and presumably for many centuries afterwards, this mountainous belt still constituted a barrier to communication which could have served to prevent linguistic innovations in the southwest from diffusing northward – except at its extreme eastern and western ends. Thus the links between southwest and midland in the Ashokan inscriptions (**6.13A**) could reflect contact through the western end of the mountainous region, and the links between southwest and east (Figure 5.4) through the eastern end. The political importance of this barrier is emphasized by Spate:

> . . . the Narbada–Chota Nagpur line . . . has been easily the most persistent internal boundary in India . . . The Aryans infiltrated beyond the Narbada, but except on the lavas of Maharashtra the 'Southland', *Dakshinapatha* or the Deccan, is still mainly Dravidian; in the E[ast], Dravidian (and other non-Aryan) languages extend farther N[orthward] . . .
>
> Spate 1954: 146–7, reprinted with permission from O. Spate, *India and Pakistan: A General and Regional Geography*, London: Methuen & Co. Ltd., 1954)

Figure 6.3 shows that the Vindhyan complex also forms a partial barrier between the midland and the east. Several early textual references also attest to the separateness of the eastern region. One is a passage in the śatapatha-Brāhmaṇa quoted by the grammarian Patanjali (*c.*100 BCE), which relates that the Asuras were destroyed because of their substandard dialect, in which **r** is replaced by **l**. This is usually interpreted to indicate that the speech of these Asuras was a Prakrit of eastern India (Thapar 1973: 411; Deshpande 1978, 1979b: 105, note 1; see also **6.14B**).[36] Another relevant passage, from the same source, describes the need of the king of the Videhas to purify with fire the land beyond the River Sadanira (OIA **sadānīra**), which had not yet been sanctified by the god Agni for habitation by Brahmans (*ŚB* I, 4, 1, 10, quoted in Thapar 1973: 419). The Sadanira,[37] now known as the Gaṇḍak (in India) or Gaṇḍakī (in Nepal), enters India near the Bihar–Uttar Pradesh border, and joins the Ganga downstream from the Ganga–Yamuna confluence. Thapar suggests that this may refer to the "boundary of Aryan [i.e. Brahman?] control in the Ganges valley" in the early Mauryan period (ibid.).

A further piece of evidence consists of the puranic accounts of the locations of various Indo-Aryan-speaking groups (frequently referred to as "tribes"). This subject has been treated by Romila Thapar, who points out that while the chronology of the puranic genealogies is not trustworthy, because of the long time gap between the original events and their recording, "they [the genealogies] can be examined as records of a general pattern of settlements and migrations" (1978: 242). As stated, for example, in the *Vishnupurana* (OIA **viṣnupurāṇa**), there were two

main branches of the ancient Indo-Aryan lineage founded respectively by Puru and Yadu, the sons of Yayāti. The Pūrus or descendants of Puru were said to be concentrated in the areas of the Indo-Gangetic divide, including the now-defunct River Sarasvati, and the Ganga–Yamuna doab (the area of Witzel's midland dialect, see **6.16**). The Yādavas, on the other hand, are spread out over a much wider area, including the Aravalli region, Malwa, Gujarat, the Narbada Valley, the northern Deccan, and the eastern Ganga Valley (Thapar 1978: 243).

Thapar has divided the Pūru genealogy into three phases. In the first phase, the Pūrus are depicted as dwelling on the banks of the Sarasvati in northern Panjab, while in the second phase the lineage divides into segments, each controlling a section of territory. In the third stage, after further divisions, new kingdoms have been established by conquest over a wider area, including some territories formerly under Yādava control, such as those of the Cedi, Vatsa, Karūṣa, Magadha, and possibly Matsya. This stage is supposed to have been reached by the time of the Mahabharata war, or sometime in the first half of the first millennium BCE (Thapar 1978: 243–9).[38] The only areas in which the Yadavas seem to have survived the expansion of the Pūrus are Gujarat, the Narmada Valley, the northern Deccan, and the east. These are all areas which lie in the territory of Grierson's "outer group."

Though according to the tradition, Purus and Yadavas belong to the same original lineage, there are several clues to indicate that the Yadavas were regarded as linguistically (and otherwise) mixed: "The inhabitants of Ānarta and Surāṣṭra are described as belonging to the *sankīrṇa jātis* [mixed or hybrid castes]. The Yādava kinship system shows traces of a matrilineal structure which is alien to the Indo-European kinship system…It is also curious that Gujarat is listed as one of the *pañca-drāviḍa* lands in a late Buddhist text" (Thapar 1978: 260–1). And note that the word Yadava (OIA **yādava-**) itself may be of Dravidian derivation (see **3.22A1**).

Some scholars (e.g. Witzel 1995a: 89–90) have criticized the use of puranic accounts as evidence for earlier periods, given that the Puranas date from the first millennium CE. Clearly these texts record traditions of uncertain age, possibly with many anachronisms (see Rocher 1986). On the other hand, it is worthwhile pointing out that at least the puranic account does not contradict the conclusions which have been reached on the basis of the linguistic evidence. The puranic genealogies depict the Purus, located in Aryavarta, as the true keepers of the **ārya** tradition, in opposition to other groups located outside that area. Though this account does not deal with the unity, or lack of unity, among those groups labeled as 'Yadavas', and equally does not address the question of how 'Yadava' groups viewed their relations with each other or with other groups, nevertheless the ideological antagonism of the people of Aryavarta toward those outside it emerges clearly.

As noted earlier, the Mahabharata epic belongs to the area of Aryavarta; the battle described in it occurred at Kurukshetra, northwest of Delhi. The Ramayana (OIA **rāmāyaṇa**), on the other hand, is set in the eastern kingdom of Kosala and

the Vindhyan forests, and its narrative probably reflects "an exaggerated version of local conflicts, occurring between expanding kingdoms of the [eastern] Ganges Plain and the less sedentary societies of the Vindhyan region" (Thapar 2002: 103). In other words, the Ramayana, in contrast to the Mahabharata, belongs to the region of outer Indo-Aryan. The original version is generally dated in the mid-first millennium BCE, though the bardic fragments from which it was assembled may be older. Though its language is more polished than that of the Mahabharata, and though its social concepts are related to a later historical period, it is traditionally believed to be the earlier of the two (op. cit.: 102–3).

6.32. Textual–archaeological correlations

Thapar (1978) notes the possible correlation between the reported territory of the Purus and the find spots of the pottery known as Painted Grey Ware (PGW), a matter raised initially by B. B. Lal (1954). It has long been known that PGW has been found at many of the sites mentioned in the Epic literature. According to V. N. Misra, "A number of sites yielding this ware like Bairat, Panipat, Purana Qila, Hastinapura, Ahicchatra and Kampil figure prominently in the story of the *Mahabharata* epic and it is therefore believed that the people of the *Mahabharata* were the same as the PGW people, and they represent the second wave of the Aryans" (Misra 2001: 521); "the culture is dated by radiocarbon to 1000–600 BC which again is in sharp contrast to the Hindu belief of the *Mahabharata* battle having ended before 3000 B.C." (ibid.). The Allchins had earlier stated that "the grey ware 'culture' occupied almost the same area as that postulated for the late Vedic settlement" (1982: 317). Witzel's study of Vedic dialects (see **6.16**) also points to this region as one of two major dialect areas at this period, and he also notes a close correspondence between the area of the Yajur-Veda **saṃhita**s and the PGW sites (Witzel 1989: 247). It is noteworthy that PGW is not found in any significant amount to the east of the Ganga–Yamuna doab.

Thapar (1975) has also suggested a possible correlation between the areas attributed to the Yadava-related groups and the sites where black-and-red ware (BRW) has been found. This is a somewhat more elusive connection, inasmuch as the term 'black-and-red (burnished) ware' is used to describe pottery finds in many different areas and at many different periods. As V. N. Misra points out, "…black-and-red ware as a ceramic is found almost all over the country from the Harappa culture in Gujarat to Megalithic culture in south India" (2001: 519). And in fact, BRW is found below PGW in a number of midland sites such as Noh and Atranjikhera, in the period 1450–1200 BCE (ibid.). On the other hand, there is a pattern which appears to be shared by the chalcolithic cultures of the eastern Ganga and those of the Narmada, in which BRW is found from the mid-second millennium BCE up until the mid-first millennium BCE, when it is replaced (at least partially) by the northern black polished ware (NBPW or NBP ware).[39]

While these various finds of BRW do not necessarily belong to the same tradition, the chronology of the NBP finds in the more southerly sites may be

significant in terms of the connection between eastern and southwestern Indo-Aryan. NBP ware is found, starting around 600 BCE, in many sites in North India, but mainly in the Gangetic plains in Bihar and eastern Uttar Pradesh. It is the characteristic pottery associated with the middle and lower Ganga valley in the period preceding the second urbanization of India in the sixth century BCE (Misra 2001: 521). Though classified as 'northern', finds of NBP in some southern sites (e.g. Kayatha, Ujjain) are as early as some finds in the Ganga valley, and are considered part of the 'early NBP phase' by Erdosy (Allchin 1995: 103).[40] This would suggest that the sociolinguistic communication networks linking the Ganga with the Narmada were as effective as those which linked the different parts of the Ganga system itself – despite the vastly greater difficulty of movement between the two river systems. These southern NBP finds are generally believed to represent imports from the Ganga region; if they came from the eastern Ganga, where NBP is most plentiful, that might tie in with the linguistic similarities between the two regions. Note that there is no attempt here to suggest a direct connection between pottery styles and language, but rather to use similarities in pottery (or other artifacts) as an indication of interregional linkages which might point to the existence of linguistic connections.

6.33. Archaeology and language movement

F. R. Allchin (1995: 41–53) has discussed the presence of 'foreign' elements in various sites which seem to imply a gradual movement of Central Asian elements successively into Afghanistan, Panjab, Sindh, and Haryana, and other archaeologists have offered similar accounts (see for example Erdosy 1995b). On the other hand, a number of students of prehistory have pointed out that there is no evidence for any substantial immigration into South Asia during this period. K. A. R. Kennedy states: "All prehistoric remains recovered thus far from the Indian subcontinent are phenotypically identifiable as ancient South Asians…[and] their biological continuity with living peoples of India, Pakistan, Sri Lanka and the border regions is well established in time and space" (1995: 60). J. Shaffer has repeatedly pointed out that there is no archaeological evidence for detectable population movement which can be linked to the introduction of the Indo-Aryan languages into South Asia (e.g. Shaffer 1984, 1986: 230, 1992).[41] While the issue that concerns the present chapter is the movements of Indo-Aryan speech communities *within* South Asia rather than the problem of how they got there in the first place, the point is an important one. Can a language move without population movements? Certainly the Spanish language moved into South and Central America without a proportionately large influx of population, but there is no evidence for this sort of Aryan elite-dominated situation in South Asia, in spite of early theories of an Indo-Aryan 'invasion'. On the other hand, there can be little doubt that an OIA speech community existed in the Panjab from the mid-second millennium BCE if not earlier, and that the language – if not the speakers – came from outside South Asia (see Erdosy 1995a; Witzel 1995a;). This is a genuine

dilemma for linguistics and for archaeology which is not limited to South Asia, and which will only be solved, if it is indeed solvable, by joint linguistic and archaeological work. In the meantime, it is the responsibility of historical linguists (or linguistic archaeologists) to make every use of the tools available to them, including the insights derived from contemporary sociolinguistic research (**1.4C**), in attempting to construct realistic models of ancient speech communities. For Old Indo-Aryan, this task is undertaken in **6.42**.

6.4. Conclusions

6.41. Preliminaries

Among students of OIA, the generally accepted picture of the expansion of the Indo-Aryan speech community within the subcontinent reflects the history of the major OIA texts, that is, from northern Panjab into the Ganga–Yamuna doab, and thence continuing along the Ganga Valley to Bihar and ultimately to Bengal, with the southern tier of languages presumably resulting from secondary southward movements from this main stream. This view is expressed by Masica, who distinguishes between the 'physical extension' of Indo-Aryan-speaking settlements and the 'Aryanization' of indigenous populations.[42] He describes the situation as follows:

> Indo-Aryan settlement, with attendant Aryanization of the first type, apparently proceeded piecemeal via the northern Punjab first into and then down the Ganges valley. Bihar was Aryanized in some sense by the time of the Buddha...Bengal was not Aryanized until the Gupta period (fourth century AD)...or even later...Orissa remained outside the Aryan pale, at least in the first sense, as late as the seventh century AD.
>
> (Masica 1991: 44–5, reprinted with permission from C. A. Masica, *The Indo-Aryan Languages*, Cambridge: Cambridge University Press, 1991)

According to this view, the southwestern Indo-Aryan languages are thought to be rather late arrivals; in Masica's words, "Some form of Indo-Aryan thrust southward – on the west side of the peninsula, not the east – ...around the eighth century BC" (ibid.). This notion, that Indo-Aryan first spread into the Deccan as the result of a southward movement from the mainstream in the eighth century BCE, is based, among other (?) things, on the finds of NBP ware in chalcolithic sites in the Deccan (Masica 1991: 45, citing Thapar 1966: 25–6). I know of no direct evidence for such a movement, and as I have suggested earlier, the presence of NBP ware in this area can perhaps be equally well explained by diffusion from the east. In any case, the data presented in Chapters 5 and 6 strongly suggest that Indo-Aryan languages were spoken in the Deccan well before this time (see above).

On the basis of the linguistic and other data presented in Chapter 5 and in the earlier sections of this chapter, it seems that this account, while it may represent the history of the main bearers of the Vedic culture and its literature, fails to tell the whole story of the prehistory of the Indo-Aryan languages. Following is a brief review of that evidence: linguistic, archaeological, and textual.

Linguistic evidence. The principal linguistic arguments are the following:

(1) A number of major linguistic innovations are shared between the western IA languages Marathi–Konkani and the eastern IA languages Bangla, Ahom (Assamese), the "Bihari" languages Bhojpuri–Maithili–Magahi, and Oriya. These innovations, which include phonological and morphological changes, are distributed in such a way that it is difficult to make a case for their occurring independently. One of these innovations, the change of OIA **ṛ** to **a**, probably goes back to Vedic times. While the other innovations may be later, it is still necessary to assume that the above-named languages (the outer IA languages) were at some time sufficiently separate from the remaining languages, and sufficiently linked with each other sociolinguistically, that these innovations were restricted to the outer group. See **5.3** and Figure 5.4.
(2) The dialects of the Ashokan inscriptions of the mid-third century BCE show that in most features the eastern and southwestern areas show more similarity to each other than to the northwestern and midland areas (**6.15**). The Vedic dialects as reconstructed by Witzel (1989) show a similar division (**6.16**).
(3) Avadhi, the language of Avadh (formerly Oudh), shows verbal forms which combine features which are characteristic of both inner and outer Indo-Aryan, and are not elsewhere found together. This agrees with Grierson's characterization of Avadh as a transitional area, and also coincides with Witzel's transitional Kosala dialect of Vedic. This can easily be explained on the assumption that Avadh is the region where the outer and inner forms of Indo-Aryan came together after centuries of separation by the Vindhyan complex (see **6.31**).
(4) Oriya, though it shares major features of the Eastern languages, has similarities to Marathi which deserve further investigation, and which may indicate that Oriya was originally in an area of overlap between Marathi–Konkani and the remaining Eastern languages within the outer group.[43]

Archaeological evidence

(1) In addition to the evidence of Central Asian traits in the Indo-Iranian borderlands in the late third–early second millennia (Allchin 1995: 48–9, see **6.33**; Erdosy 1995b:10), there is further evidence of intrusive elements in sites in Sindh, Panjab, and Haryana, 1700–1200 BCE (Allchin 1995: 49, see **6.33**).
(2) Allchin mentions the existence of an ancient route which "leaves lower Sind and crosses between the desert and the marshes of Cutch, re-emerging in southern Rajasthan and the fertile Malwa plateau" (Allchin and Allchin 1982: 249).

(3) Artifacts of "Iranian" affinity are found in sites in Malwa and the Deccan in the mid-to-late second millennium BCE (see **6.33**).
(4) Possibly, the findings of black-and-red ware in sites on the Narmada, Tapi, and lower Ganga may be relatable to the locations of speakers of outer Indo-Aryan (see **6.32** and further discussion below).

Textual evidence. This can only be indirect, since as noted above the OIA texts relate primarily to the mainstream movement from northern Panjab eastward. The following points may be mentioned:

(1) There is ample evidence for the presence of groups speaking aberrant forms of Indo-Aryan which were considered to be outside of the Vedic society; in fact, the Dasas (OIA **dāsa**), who were previously thought to speak Dravidian or other non-Indo-Aryan languages, may also have been Indo-Aryan speakers (Parpola 1988; Erdosy 1995b: 15).
(2) The **mleccha** territories listed in the śatapaṭha Brāhmaṇa include not only the southwestern areas of Saurashtra, Gujarat, and Maharashtra, but also the eastern areas of modern-day Bihar and Bengal (see above) – in spite of the fact that Indo-Aryan languages were spoken in these areas.

6.42. A reconstruction of the prehistory of outer Indo-Aryan

While the evidence listed earlier is not sufficient to allow a detailed history of the 'non-Vedic' part of the OIA speech community, some sort of outline can be sketched. The following account is divided arbitrarily into four phases: (1) the Indus Valley, (2) expansion from Sindh into peninsular India, (3) further expansion into eastern India, (4) the reconnection of inner and outer Indo-Aryan in the region of Avadh. These movements are summarized in Figure 6.4.

(*1*) *Indo-Aryan in the Indus Valley*: By 1500 BCE, when the first hymns of the Rigveda are believed to have been composed, that portion of the Indo-Aryan speech community which was associated with the OIA texts was located in the upper Indus Valley. The OIA language was already distinct from Old Iranian, though there may have been intermittent contact between the two, as indicated by parallel textual passages. The texts show a strong emphasis on animal herding, though agriculture was also practiced, presumably by Indo-Aryan speakers as well as indigenous groups. Thus the OIA speech community was already embedded in a multilingual and multicultural society. Those OIA speakers who practiced herding probably did not have fixed abodes, and thus this speech community was different in more than one way from those which modern sociolinguists have studied.

Given the archaeological evidence for intrusive Central Asian elements on the lower Indus (see earlier text), it can only be assumed that OIA speakers also occupied this area by the end of the second millennium BCE. Since there is no mention of

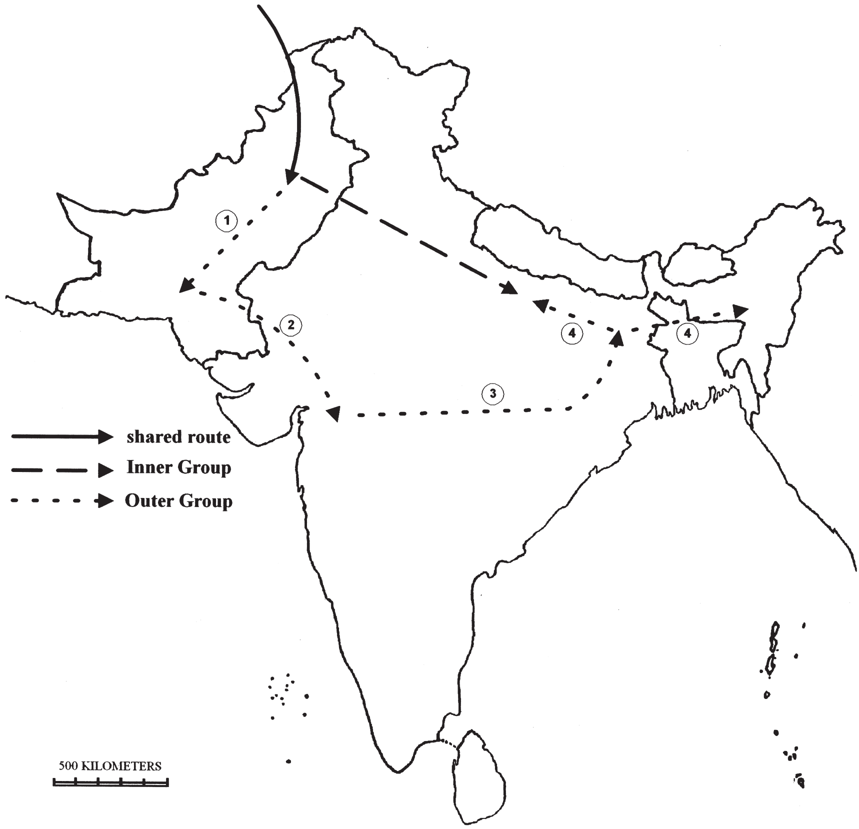


Figure 6.4 Proposed movement of Indo-Aryan languages 2000–500 BCE.

Sindh in OIA texts until the late/post-Vedic period, it must also be assumed that the speakers of OIA in Sindh were not considered to belong to the Vedic society. Thus we posit a geographically expanding speech community, and it follows reasonably that as the OIA 'mainstream' society expanded eastward across the Indo-Gangetic divide and into the Ganga–Yamuna doab by the end of the Rigvedic period (*c.*800 BCE), its counterpart in Sindh probably did the same, following the route mentioned here, leading to Malwa, Gujarat, and the Deccan.[44] In fact, this eastward movement from Sindh may have started some centuries earlier, given the evidence of intrusive northwestern ('Iranian') elements in Malwa in the mid-second millennium BCE mentioned above. (See **9.4** for evidence favoring this earlier dating.) These finds might be considered to represent the earliest phase of this southern expansion, which can be presumed to have ended before the large-scale abandonment of sites in Sindh around the turn of the first millennium BCE (Allchin 1995: 36).[45]

The OIA language in the Indus Valley clearly showed from the earliest times the usual kinds of regional, social, and contextual variation (see **1.4A, 2.8**, also Salomon 1995). The very existence of sacred texts and the extreme measures used to preserve them orally is an indication of an awareness that linguistic change was taking place.[46] Given the geographical extent of the community, regional variation was also inevitable, and the texts give some hints of this also (see Witzel 1989). One important example of regional variation is relevant for the present discussion: the change of OIA **ṛ** to **a/i/u** in different areas had already begun by the end of the Rigvedic period (see **6.12B, 5.22A**). Assuming that this change took place during the Indus Valley phase, the present distribution of the reflexes of **ṛ** in NIA would indicate that **i** was the northern representation and **a** the southern reflex. The situation is somewhat more complex, however, since a number of words which contained OIA **ṛ** are found with the same vowel (whether **a** or **i** or **u**) in all the MIA and NIA languages (see **6.12B**). Such words may reflect lexical diffusion in the earliest period of Indo-Aryan, that is, cases where the variant found in one region, or in one social dialect, was for some reason generalized to the entire speech community. This would imply that the Indo-Aryan speech community of that time, though already showing regional differentiation, was still compact and homogeneous enough for innovations to travel across it. This would be compatible with a situation in which a substantial amount of the population was in movement, with perhaps some segments remaining relatively stationary.

Speakers of some Dravidian language(s) are believed to have occupied parts of Sindh from about the twelfth century BCE, if not earlier, as well as parts of Maharashtra and the Deccan (Chapter 9). (See **3.22A** and **10.21** for possible earlier presence of Dravidians in the Indus Valley.) Thus speakers of outer Indo-Aryan most likely had greater contact with Dravidian speakers than did those who spoke inner Indo-Aryan, and in fact the former may have served as intermediaries between Dravidian speakers and inner Indo-Aryan speakers in transmitting the Dravidian loanwords found in OIA texts.

(*2*) *Expansion into the peninsula.* In Central India and the Deccan, there is widespread evidence of settlements going back to the early second millennium BCE or earlier, and Allchin suggests that the rise of cities in these regions may have involved "indigenous populations contributing their own important element, alongside the possible stimulus produced by movements of peoples originating in the north or northwest" (Allchin 1995: 134). The cities in question include those belonging to certain of the Mahajanapadas (population centers) which were known to exist at the time of the Buddha, for example, Ceti/Cedi (which probably included the cities of Tripuri near modern Jabalpur in the Narmada valley, and Eran near present-day Sagar), Avanti (whose probable capital was Ujjain, and which included a second larger city, Vidisa), and Asmaka (possibly the region of Paithan–Pratisthana and Ter–Tagara in the Godavari region) (ibid.). It may of course be difficult to find evidence of the separate contributions by local peoples and newcomers to the development of these cities, though all the sites mentioned

can be dated within the period suggested here for the expansion of outer Indo-Aryan into these regions (see above).

(*3*) *Expansion into eastern India*. The linguistic evidence for shared innovations in outer Indo-Aryan implies a further expansion from Central India into the eastern region. It was suggested in **6.31** that this eastward expansion progressed along the southern boundary of the Vindhya complex into what is now the new state of Jharkhand, and thence northward into the Ganga valley. The situation was doubtless more complicated. The Vindhya complex was never an absolute barrier to movement, and there has been traffic across it between the Ganga and Narmada valleys from early times (Allchin 1995: 134). Part of this movement is reflected in the puranic accounts of Yadava peoples like the Cedi, Matsya, Vatsa, and Karusa, which were claimed as conquests by kings of the Puru lineage (**6.31**). It is likely that there was movement in both directions, involving traders, military expeditions, migrant groups, etc., along the valleys of the Chambal, Sind, Betwa, Ken, Son, and other rivers. While the linguistic peculiarities of southern speakers who migrated into the areas already settled by speakers of inner Indo-Aryan would not have survived, the evidence suggests that those who moved farther east were predominantly speakers of outer Indo-Aryan. It is possible that this eastward movement was encouraged by the fact that the areas of the Ganga–Yamuna doab and the central Ganga had been settled earlier.

It can be presumed that the eastern region was an area of outer Indo-Aryan speech by the time of Ashoka, because of the linguistic similarities between the eastern and southwestern Ashokan inscriptions. In fact, this was most likely the case from the beginning of the Mauryan Empire in the fourth century BCE, if not earlier. The linguistic features which distinguish the modern outer and inner languages, apart from the late Rigvedic change **ṛ** → **a**, must have developed and spread between the time of the movement from Sindh into the peninsula and the Mauryan period, though some changes may have diffused into the eastern region later (see **6.2**). The reflexes of the **-tavya** gerundive (**5.21A**) diffused northward into inner Indo-Aryan, at least sufficiently to become acceptable to grammarians like Panini (*c.*500 BCE).[47] We have at present no means of dating these linguistic changes. The spread of the **-l-** past from west to east could have taken place at any time during this period, though it is more likely to have occurred before the changes undergone by the eastern group of languages (**6.14**). As noted earlier, the late attestation of the **-l-** past creates a problem which is as yet unresolved (see **6.2**).

(*4*) *Reconnection of inner and outer branches.* The final phase of Indo-Aryan diffusion, which is perhaps still ongoing in some areas (see Masica 1991: 45), involves further spread and consolidation in the areas into which the languages diffused in the previous stage. The two branches of Indo-Aryan came together in the central Ganga Valley – presumably in the region of Avadhi, as evidenced by the presence in that area of mixed forms of the future and the gerundive, and sporadic occurrence of **-l-** past forms.[48]

It must be assumed that, both during and after this last phase – and possibly earlier as well – other linguistic features diffused across various regions of

Indo-Aryan, so as to produce the seemingly chaotic patterns shown in Figure 5.2. No other features, however, cover as wide a range as those which have been used here to define the southern or outer branch of Indo-Aryan; nor is there any case of multiple innovations showing as much overlap in their distributions (see Figure 5.4). This is the primary, and ultimate, justification for using these features to infer the existence of a major early division in Indo-Aryan, in spite of the (apparent) differences in the chronological development of these features.

It should be emphasized that, although the Grierson hypothesis has often been linked to the belief in two or more successive 'waves' of Indo-Aryan speakers migrating into South Asia, the evidence discussed in Chapters 5 and 6 does not require any such assumption. All of the movements which have been posited earlier can be accounted for by assuming a long, slow influx of Central Asian herding peoples moving into the Indo-Iranian borderlands, to the Panjab, and thence on the one hand eastwards to the Ganga, and on the other hand down the Indus to the Deccan and further east. Such an assumption of course does not resolve the question of why so few traces of such movement have been found (see **6.33**), nor the question as to why and how the Vedic culture developed and was cultivated in only one part of the Indo-Aryan speech community. As yet, we do not have enough information to answer these questions, regardless of which assumptions we make.

6.5. Summary

Following up the evidence of Chapter 5, which shows linguistic links between the eastern and southwestern regions of NIA conflicting with the accepted view of NIA subgrouping, this chapter examines the linguistic history of MIA as well as the prehistory and social history of the region in order to propose a resolution. An examination of the geographical distribution of linguistic traits in the Ashokan inscriptions (third century BCE) indicates that the majority of innovations appearing in the inscriptions show agreement between the east and southwest. This evidence is summarized in **1.5**, where it is pointed out that the major dialectal divisions in MIA are between the northwest–center and the rest (southwest and east), with the midland area sharing individual innovations with the three adjacent areas (northwest, southwest, and east). Thus, the MIA data agree with the data from the modern languages presented in Chapter 5. The evidence of the Vedic dialects shows a similar division (**6.16**). In addition, the eastern languages are united by a group of innovations which set them off from the remaining languages; these innovations, which probably occurred after those innovations shared between southwest and east, appear in the mid-third century BCE, implying that the common southwest–east changes probably occurred during or before the first half of the first millennium BCE.

Summarizing all the linguistic evidence to date, **6.2** notes that all of the evidence of shared innovations indicates an early dialect division between the languages of the northwest–center and the southwest–east. Chronology, however, is a problem: while the evidence of the change of **ṛ** to **a** (SW–E) and **i** (N–C) suggests that this dialect division could have existed during Vedic times, the past forms in **-l-** do not

appear in texts until the eighth century CE. A possible social mechanism is proposed for this long time gap. In **6.21** it is noted that some of the structural similarities shared by eastern languages like Bangla and southwestern languages like Marathi are also shared by Dravidian languages, indicating the likelihood of early influence of Dravidian (and perhaps other indigenous languages) on outer Indo-Aryan.

The regional division represented by the N–C and SW–E languages corresponds to a longstanding historical division between Aryavarta, the land of Hindu orthodoxy, and the regions known as *mleccha-deša*s or barbarian lands, where it was understood that non-Aryan languages were spoken and Hindu rituals were not observed. Perhaps significantly, this distinction continued to appear in the literature even long after communications had been established between the central and peripheral areas. These two areas are geographically separated by the Vindhyan complex, a continuous chain of mountains, hills, and plateaus which stretches across central India. Accounts of the branches of the Yādava clan (a branch of the ancient Indo-Aryan lineage which was reputed to have been tainted by the adoption of local customs) place these people mainly to the south or east of the Vindhyan complex (**6.31**). According to the Allchins and other archaeologists, there is an apparent link between the find-sites of PGW and locations associated with OIA, especially the Mahabharata (**6.32**); PGW is not found in the area now occupied by the outer languages.

The OIA literary tradition reflects a movement of Indo-Aryan speakers from the upper Indus to the Ganga–Yamuna doab, with later movements from the doab toward the east and south. The archaeological record shows, in addition, evidence of movements southward on the Indus and eastward from the lower Indus. Following a brief summary of the linguistic, archaeological, and textual evidence (**6.41**), Section **6.42** proposes a four-stage sequence for the movement of the outer Indo-Aryan languages into the subcontinent:

(1) southward expansion from the northern Indus into Sindh by the middle Rigvedic period;
(2) eastward movement from Sindh into Malwa and the Deccan, paralleling the movement of inner Indo-Aryan from Panjab into the Ganga–Yamuna doab;
(3) continued eastward movement into eastern India, followed by northern movement to the lower Ganga before the time of the Buddha and the Mauryan kings (fourth century BCE);
(4) consolidation of the eastern group of languages and their reconnection with the inner languages in the region of Avadh.

Notes

1 "On peut donc raisonnablement attribuer à l'influence de l'original de Pataliputra quelques-uns des traits qui, a Kalsi par exemple, paraissent aberrants par rapport aux normes du pk. 'administratif' local…" (Caillat 1989b: 416) – in contrast with the west and northwest, of which she says, "…le ouest et le nord-ouest conservent une

autonomie plus grande: les formes de m.i. qui s'y trouvent employées apparaissent comme des alternantes assez nettement marquées par rapport à celles qui se rencontrent ailleurs dans le royaume" (1989b: 419). Caillat points out, on the other hand, that the inscriptions of the central area do not appear to be slavish copies of eastern originals (1989b: 431).

2 The exceptions to these statements mostly involve words which show the same vowel throughout NIA, such as OIA **dṛśyate** 'is seen' (CDIAL 6516) which shows **i** throughout MIA and NIA, OIA **nṛtta** 'dancing' (CDIAL 7580) with **a** throughout, OIA **pṛcchati** 'asks' (CDIAL 8532) with **u** throughout.

3 A single exception is the **i** in **kiṭanātā** 'gratitude' (← OIA **kṛtajñatā**) at Kalsi (G. **katañññatā**, Sh. **kiṭrañata**, missing in Dh. and J.).

4 The form **sarva** 'all' in Girnar, which alternates with **savva**, may be regarded as a straight **tatsama** borrowing from Sanskrit. (Modern elite Marathi uses the same form.)

5 The Marathi **s** and **ś** (from whatever OIA source: **kṣ**, **ch**, sibilant, or other) were earlier differentiated by phonological environment, with **ś** occurring only before **i y e**: **śikṇe** 'to learn', **ś(y)ām** 'Shyam' (proper name), **śe** '100', **khisā** 'pocket', etc. With the later loss of **y** the **ś** appeared in new environments, e.g. **khiśāt** 'in the pocket' (← **khis-** [stem] **-yā-** [oblique] **-t** [in], cf. **hā** 'this one', **(h)yā-t** 'in this one').

6 Michael Witzel has pointed out to me that Nepali has both **kh** (e.g. in **khet** 'field' ← OIA **kṣetra**) and **ch** (e.g. in **chetri** 'Chetri (caste)' ← OIA **kṣatriya**).

7 Chatterji, on the assumption that vowel-insertion was the normal outcome in Bangla, inferred that the eastern future forms in **b** (← **v** ← **vv**) must have been the result of the influence of some western dialect. In the light of present evidence, such an inference seems unwarranted. (See **5.21A**.) Note that in modern Hindi, **Cy** clusters in tatsama words are occasionally pronounced with vowel-insertion, e.g. **samasiyā** for **samasyā** 'problem'.

8 The three-way contrast cannot be fully illustrated in any position from the Ashokan materials, since **ṇ** does not occur initially and there are no examples of medial **-ñ-** or**-nn-**.

9 The distinction has been reintroduced in **tatsama**s, e.g. Hindi **kāraṇ** 'reason', though this **ṇ** is usually rendered as dental /**n**/ in normal speech.

10 MIA frequently shows initial (orthographic) **ṇ**, and it is not clear whether this represents phonetic [ṇ] or [n]. In any case, the evidence for the existence of the **ṇ**–**n** contrast in MIA is equivocal. (See Bloch 1965: 78–9 [original pagination].) Mehendale's survey of Prakrit inscriptions (1948) seems to indicate that while **ñ** (from whatever source) merged with **n** everywhere, the distinction between **ṇ** and **n** survived at least partially in the central (p. 158) and eastern (p. 88) areas. There is no inscriptional evidence from the northwest, and the evidence from the western area is conflicting: according to Mehendale, the change **ṇ** → **n** is "universal" in western inscriptions (1948: 64), while at the same time the "cerebralization" of **n** (i.e. **n** → **ṇ**) is frequent, and even universal in some inscriptions (71–4; see also 212, 222–3, 273).

11 See further discussion of this passage in Deshpande (1979b: 4).

12 Pa = Pali, Pk = Prakrit, Dr = Dardic, K = Kashmiri, S = Sindhi, LP = Lahnda-Panjabi, PN = West Pahari–Garhwali–Kumaoni–Nepali, A = Assamese, B = Bangla, O = Oriya, Bi = "Bihari"–Bhojpuri–Maithili, H = Hindi, G = Gujarati, M = Marathi–Konkani.

13 Scores are calculated by counting ∅ for each **r-**, 1 for each **l-** (or **n** ← **l**), and 0.5 for each case of ~, then dividing the total by the number of items in the column.

14 Items 1–13 are those which show only **r-** in OIA; the remaining items have OIA forms with both **r** and **l**.

15 Masica accepts the standard account:

> Another Magadhan feature, the universal substitution of *l* for *r*, has left only a few traces (in terms of forms with *l* in place of Sanskrit *r*) in the daughter languages, a few more in Assamese than in Bengali (As. *lāi*, Beng. Or. *rāi* 'black

mustard-seed' (< *rājikā*). The Magadhan forms with *l* (which are well-attested in inscriptions as well as in the stylized language of the drama and in the descriptions and prescriptions of the grammarians throughout the MIA period) were apparently overwhelmed later on by Midland and Sanskrit influences.

(1991: 186, reprinted with permission from C. A. Masica, *The Indo-Aryan languages*, Cambridge: Cambridge University Press, 1991)

16 On the other hand, Colette Caillat (1989b: 419) suggests that these cases of western **r** for **l** may be the result of Iranian influence.

17 Some close parallels can be found between this situation and the non-Brahman movement in Tamilnadu in the 1960s and 1970s; see for example Hardgrave (1965) and Irschick (1969).

18 This word is generally thought to be of the same origin as the Sanskrit name **Rukmiṇī**, probably connected with OIA **rukma** 'what is bright or radiant' and thus going back to a root with PIE ***l** (cf. Lat. **lūx** 'light' **lūmen** 'light', Gr. **leukós** 'light, white', OHG. **licht**, etc.).

19 Pali, the principal language of the Buddhist scriptures, has a number of cases of **l** for OIA **r** (see the above list). As to the regional affiliation of Pali, opinion is divided among the earlier scholars. The Buddhist tradition in Srilanka claims that Pali is Magadhi, and that the Pali Tipitaka is composed in the language used by the Buddha himself. This has been rejected by a number of scholars, primarily because Pali does not show the characteristic phonological and morphological features of Magadhi (including the wholesale change of **r** → **l**), and a number of persuasive arguments have been given for considering Pali to be a western dialect. (Geiger notes that "Pali however retains the **r** (its change into **l** is indeed frequent but not the rule)" (1938: 3).) Others have suggested the Kalinga country (roughly modern Orissa). In the absence of a consensus, Geiger came to the conclusion that Pali, though to some extent a composite or **koiné**, was probably based on the language of the Buddha (Geiger 1978: 2–5). Since Pali exhibits both the retention of some cases of PIE ***l** which had been replaced by **r** in OIA, as well as the sporadic change of earlier **r** → **l**, it is what would be expected according to the above mentioned argument, that is, a composite dialect. According to Geiger, "the language of the Jaina-suttas [which] is called Ardha-Māgadhī, i.e. 'half-Magadhi'…differs from Māgadhī proper on similar points as Pāli": for example, it avoids the wholesale change of **r** to **l**, and shows the **-o** nominative of nouns rather than the Magadhan **-e** (Geiger 1938: 5). Possibly the term 'Ardhamagadhi' indicates that this literary Prakrit is a compromise between the literary Magadhi and the speech of the western regions where the Jain texts are believed to have been composed.

20 The inscription of Rupnath, in modern Madhya Pradesh, seems to have no discernible system in its use of **r** and **l** (Bloch 1950: 47).

21 In this context, the substitution of **l** for **r** can be likened to the (formerly) common stereotype of speakers from Brooklyn, New York, who supposedly merged the vowels of **verse** and **voice**, **Earl** and **oil**. In fact, it was mainly the **-er** words (**third, shirt, curl, worm**) in which the diphthong [əɪ] (phonemically /ʌy/) occurred. William Labov notes,

This sound is still frequently heard in New York City…A few lower and working class respondents used this diphthong for *oil* and *voice*, as well as for *Earl* and *verse*. But this merger of word classes is rare today; middle class speakers have apparently never used /ʌy/ for the *voice* group of words, even when they used it regularly for *verse*.

(Labov 1966: 338, reprinted with permission from W. Labov, *The Social Stratification of English in New York City*, Washington, DC: Center for Applied Linguistics, 1966)

This example is parallel to the alleged merger of **r** and **l** in the eastern Prakrits in another sense also: as Labov points out, "…Brooklynese…follows the usual pattern of attributing class differences to geographic locations" (personal communication, 27 February 2003). In the Prakrit case, it seems likely that a stereotyped perception of the speech of one social group (Buddhists) was attributed to the whole eastern Indo-Aryan region.

22 In classical Magadhi, this change is restricted to the masculine singular ending of thematic nouns; in the Ashokan inscriptions (and in other inscriptions, see Mehendale 1948) it is found in the neuter singular also.

23 Parpola argues that the Magadhi Prakrit originated in the northwest, on the basis of the following features: (1) The **-e** (← **as**) nominative sg. endings in the NW and E inscriptions **(6.14A)**, (2) the occasional occurrence of **l** ← OIA **r** in the late RV, which is relatable to the change **r →l (see 6.14B)** in the E inscriptions of Ashoka, and (3) the change **s ṣ** → **ś** found in "the Prakrit of northern Sindh called Vrācaḍā" (Parpola 2002a: 234). While this possibility cannot be ruled out, the following points are relevant:

(1) The **-e** endings in the northwest are as likely to have resulted from scribal imitations of the eastern originals, given that there is no trace of this ending in northwestern NIA.
(2) I have argued, again on the basis of NIA evidence, that the wholesale change **r** → **l** in the eastern inscriptions was restricted to written language, and never actually occurred in speech **(6.14B)**.
(3) Except for the NW, all the Ashokan inscriptions show a single graphic character for the three OIA sibilants **ś ṣ s (6.12A)**. This is confirmed by the evidence of NIA, which generally shows the character transcribed as **s** (though affected in some cases by later changes, for example Marathi **s** → **ś** before **i e y**) except in the NW. This **s** is phonetically [s] in most environments, except in the modern E languages where it is most commonly [ʃ]. This appears to be the result of a (merely phonetic) change restricted to the E languages, probably *subsequent* to the more general NIA merger of the three sibilants. It is probably impossible to date the eastern change, since the spelling has never varied. Thus there is little basis for linking this change to the *phonemic merger* which occurred in a particular NW dialect, as well as elsewhere in MIA and NIA.

24 The occurrence of reflexes of the **-tavya** gerundive in the northwestern Ashokan inscriptions is difficult to explain, inasmuch as it was not present in the Rigveda and is not present in NIA in this region. This difficulty exists whether or not the "inner–outer" hypothesis is accepted.

25 Note that I have suggested in **6.14B** that the change **r** → **l**, supposedly characteristic of the Eastern inscriptions of Ashoka, probably never actually took place in the spoken language of the majority of people in this area.

26 Klaiman herself mentions a Sinhala parallel for the Bangla and Marathi complementizers (1977: 307), and a Gujarati parallel **thaki** for the postposition meaning 'from' derived from a copular verb (1977: 306) are discussed here. Other parallels include

(1) M **soḍ-un** 'apart from, except for' (lit. "having-left"), parallel to T **viṭa**, B **chāṛā** (Klaiman 1977: 304);
(2) M **pāh-un** 'compared with' (lit. "having-seen"), as in **te pāhun he phār lahān āhe** 'Compared to that one (lit. "Seeing that…") this one is very small' op. cit. 305);
(3) M **pāhije(t)** 'is (are) needed/wanted', parallel to T **vēṇṭum**, B **cāi** (*op. cit.* 308) – but cf. Hindi–Urdu **cāhiye**, with closely similar meaning and usage;
(4) parallel to the clause-final conditional markers in B and T (op. cit. 309), we find in Marathi (and also in Hindi–Urdu) a tendency to delete the clause-initial marker but to retain the (originally correlative) clause-final marker: thus M **[jar] pāus ālā**

tar jāṇār nāhi 'If it rains we won't go' (lit. "[if] rain came then going not", H-U **[agar] bāriṣ āī to nahī̃ jāẽge** "[if] rain came then not we-go";

(5) like B and T, M also generally places negative and interrogative markers at the end of the clause (op. cit. 310): M **tumhi gelā kā** 'Did you go?' ("you went Q"), **mi gele nāhi** 'I didn't go' ("I went not"), whereas in Hindi–Urdu the negative precedes the verb in unemphatic sentences: H **mãĩ nahī̃ gaī** 'I didn't go' ("I not went"); on the other hand, in colloquial Hindi–Urdu the interrogative marker often occurs sentence-finally: H **tum gae the kyā** 'Did you go?' ("you gone were Q");

(6) As Klaiman notes (op. cit. 311–12) M, like T and B, has a series of negative verbs (see Southworth 1976b);

(7) Hindi–Urdu and Nepali use the conjunctive participle of the copular verb in the sense of 'by way of': H **vah mathurā hoke āī** 'She came by way of Mathura', a usage related to that of T **iruntu**, B **theke**, M **houn**, etc.

27 Furthermore, the Bangla **bole** and the Marathi **mhaṇun** also share the meaning 'because': Marathi **payse navhte mhaṇun gelo nāhi** "money was-not **mhaṇun** I-went not" = 'Since I had no money I didn't go' (see Klaiman 1977: 307). The link between the quotative and the causal meanings can be seen in Marathi sentences like **mi damlo hoto mhaṇun ārām ghetlā** "I tired was **mhaṇun** rest took" = 'Saying/thinking "I'm tired", I took rest'. Such usages are attested in OIA; see MacDonnell's example **nāvamantavyo manuṣya iti bhūmipaḥ** "is-not-to-be-despised human-being **iti** king" = 'A king should not be despised, thinking/saying (he is a) human being' (MacDonnell 1927: 148).

28 See Chatterji, ODBL, 751ff. for the history of this form and traces of other forms in eastern IA.

29 I have suggested elsewhere (Southworth 1974: 07) that at least some of these "genitive" postpositions derive from verbal participles; for example, the Panjabi **rām dā** 'Ram's' ← MIA ***rāmā diyā** ← OIA ***rāmāya ditaḥ** 'given to Ram', on the model of constructions like Tamil **rāman-uṭaiya** 'Raman's' (literally "Raman-owned").

30 In the case of the quotative particles (Bangla **bole**, Marathi **mhaṇun,** etc.), these clearly continue the OIA **iti**, which seems to show Dravidian influence in its usage, and may even be of Dravidian origin (see **3.32**).

31 For example: **S** = [**bābu āyo** "father came"] **Comp** = [**bhanera**] **VP** = [**sapanā dekhio** "dream saw"] 'He dreamt that his father had come', **S** = [**chuṭṭī cāhincha** "leave wants"] **Comp** = [**bhanera**] **VP** = [**daftar mā āyo** "office in he-came"] 'He came to the office because he wants leave' (or '…saying he wants…'). On the other hand, Nepali seems to have other features in common with Bangla, such as the loss of grammatical gender, the rounding of short **a**, etc.; Thus Nepali may represent another transitional area between the inner and outer groups.

32 OIA *mleccha* means 'barbarian, non-Aryan, speaker of a foreign language, stammerer'. See Thapar (1973) and the Appendix to Chapter 3, item A40, for further discussion of this word.

33 "The *Baudhāyana Dharmasūtra* (1.1.32–33) gives us a clear idea of how the "Vedic Aryans" viewed the "mixed Aryans" of the outer regions:

> The inhabitants of Ānartta, of Anga, of Magadha, of Saurāṣṭra, of the Deccan, of Upavṛt, of Sind, and the Sauvīras are of mixed origin. He who has visited the countries of the Ārṭṭas, Kāraskaras, Puṇḍras, Sauvīras, Vangas, Kalingas [or] Pranūnas shall offer a Punastoma or Sarvapṛṣṭhī sacrifice [for purification].
>
> (from Deshpande 1979b: 265; see also his note 1, p. 105, and note 22, p. 107)

34 Deshpande points out that speakers of Indo-Aryan who belonged to regions other than Aryavarta considered themselves *ārya*s, not *mleccha*s, and in fact had an entirely different definition of what constituted *ārya* status (Deshpande 1979b: 48).

35 The stabilizing of what were to be the Arya-lands and the mleccha-lands took some time. In the *Ṛg Veda* the geographical focus was the *sapta-sindhu* (the Indus Valley and the Punjab) with Sarasvatī as the sacred river, but within a few centuries *ārya-varta* is located in the Ganga–Yamūnā Doāb with the Ganges becoming the sacred river. Together with the shift eastwards of 'the pure land' the northern Punjab and the trans-Indus region came to be regarded as *mleccha-deśa*

(Thapar 1978: 159)

This change had apparently taken place by the time of Yaska's Nirukta (mid-first millennium BCE). See Deshpande 1979b: 5.

36 Deshpande suggests that this might be "a hidden reference to the defeat of the Mauryas by the Brahmin Puṣyamitra śunga who may have been Patanjali's patron" (1979b: 9).

37 The river name OIA **sadā-nīra/nīrā** presumably means 'ever-flowing', the last element being generally accepted as a borrowing from Dravidian (PD***nīr** 'water' DEDR 3690; cf. EWA s.v., CDIAL 7552). Witzel points out that the modern name for this river, Gaṇḍak/Gaṇḍakī, seems to reflect the Munda word for 'water' (1995a: 105).

38 According to Misra (2001) (see **6.32**), if the dates of PGW are considered to be connected with events in the Mahabharata, this stage should probably be placed in the *second* half of the first millennium BCE.

39 "…the so-called Chalcolithic period, datable to the mid-2nd millennium BC and identifiable by the presence of Black and Red ware in the lower Ganga valley and by Ochre Coloured Pottery further west…" (Allchin 1995:76); see also op. cit. 102. The Allchins mention BRW in several sites of the middle Ganga from 1500 BCE up to the appearance of NBP ware around 500 BCE (Allchin and Allchin 1982: 320), in sites on the Narmada estuary (1982: 326), at Prakash in the Tapi Valley (1982: 327), and at Ujjain on the Narmada (ibid.).

40 NBP has also been found in sites in Maharashtra such as Paiṭhaṇ (Pratiṣṭhāna) and Ter (Tagara), though these finds are not precisely dated; see Allchin (1995: 139). Other finds outside of the Ganga or Narmada valleys, for example in coastal Orissa, Tamilnadu, and Srilanka (Allchin 1995: 142, 146–9, 174–5) appear to be later, and may be the results of maritime contacts.

41 It may be worth noting here that knowledge of the presence of Indo-Aryan speakers among the Mitanni (see **2.21**) is also entirely dependent on linguistic evidence.

42 Masica also distinguishes a second type of "Aryanization," described as "retention of non-Aryan speech in an Aryan garb and its cultivation under the tutelage of Sanskrit (or Pali)" (1991: 44), referring for example to Sanskritized forms of Kannada and Telugu.

43 This statement is tentative, as the similarities are rather fragmentary at this stage, though nevertheless suggestive:

(1) Oriya is the only eastern language to retain the contrast between **n** and **ṇ**, which is also retained in Marathi, Gujarati, Sindhi, and Panjabi (**6.14A**).While this is a case of retention, not innovation, it does appear to indicate that Oriya was separated from the other eastern languages at the time they underwent this change.

(2) Marathi and Oriya share some irregular past forms not found elsewhere, for example, **ghe-** 'take', past **ghe-t-(a)lā (5.13)**, see CDIAL 4509 ***ghṛpta**.

(3) Marathi and Oriya share certain detailed lexical peculiarities not found elsewhere, for example the final **-r** in the word **nāngar** 'plough' (**5.22D**).

44 Allchin provides an archaeologist's view of the early Indo-Aryan period in South Asia, which contrasts with the statement of Masica quoted in **6.41**:

…there was probably a long period of movement to and fro from Central Asia and the frontiers of the Indus; followed by further movements into the Indus

heartland; and by a series of later dispersals farther towards the east and south.

(Allchin 1995: 50–1, reprinted with permission from R. Allchin, *The Archaeology of Early Historic South Asia: The Emergence of Cities and States*, Cambridge: Cambridge University Press, 1995)

45 While Sindhi shows some outer group features, as noted in Chapter 5, it seems to belong primarily to the inner group. For example, it predominantly shows **i** as a reflex of OIA **ṛ** (see examples later in text). It is possible, then, that the few outer features found in Sindhi represent diffusion from Gujarat–Saurashtra, which would make Sindhi mainly an inner language with some outer influence, in contrast to Gujarati which seems to be an outer language with an inner overlay. This would imply the probability that the Sindhi language is the result of reoccupation of the region from neighboring areas of inner Indo-Aryan, following the abandonment of sites noted here.

46 Witzel suggests that even in the early Rigvedic period, the hymns of the Rigveda "…could only be appreciated properly by the more educated sections of a contemporary audience," implying that their form was markedly different from that of ordinary spoken language.

47 As noted in **5.21B**, this form occurs in Bandeli (Northwestern Madhya Pradesh) and in Rajasthani, though not in the other inner NIA languages.

48 The future indicative in early Awadhi has both **-h-** forms (← OIA **-iṣya-**) and **-b-** forms (← OIA **-tavya**) in all three persons and both numbers (Saksena 1971: 261). In modern Awadhi the **-b-** forms occur only in the first person plural in the western dialects, but in both first and second person plural in the eastern dialects (Saksena 1971: 264). A few **-l-** past forms are found in an eighteenth-century manuscript written in an eastern variety of Awadhi, though Saksena opines that "obviously they are borrowed from Bihari" (1971: 246). Perhaps "diffused" might be a better word than "borrowed." Early and modern Awadhi both have verbal nouns in **-ana** (← OIA **-anīya**) and **-aba** (← OIA **-tavya**, see **5.21A**). Sociolinguistically, the mixed future paradigms would appear to be a case of original "free" variation being resolved by grammaticalization, in this case redistributed as singular and plural forms.

7

PALAEOBOTANICAL AND ETYMOLOGICAL EVIDENCE FOR THE PREHISTORY OF SOUTH ASIAN CROP PLANTS

Contents

7.0. Introduction

This chapter brings botanical and palaeobotanical[1] information on South Asian crop plants together with etymological data on the histories of plant names, in an effort to illuminate the historical relationships between plants and human societies in South Asia. As I noted in an earlier paper (Southworth 1988), relationships between social groups and the plant world form a basic part of human culture and subsistence patterns, and are therefore important in the study of social – as well as linguistic – history.

Archaeobotanical investigations draw on botany and anthropology/archaeology to generate inferences about prehistoric interrelationships between human culture and the plant world. The principal questions which are asked in these investigations are: (1) What plants were recovered? (2) What can be said about the use of these plants? (3) What can be further inferred about the plant-use strategies and

the subsistence base of the prehistoric inhabitants of the site? (4) What can be further inferred about the habitat at the time?

The data for these investigations are made up of plant parts, including seeds – which usually means carbonized seeds, since unburned seeds generally are not preserved over long periods of time. Seeds may be carbonized in several ways: from spillage during cooking, from the use of animal dung for fuel, from the burning of straw, from naturally occurring fires, etc. While other plant parts such as straw, leaves, and husks may be of value in identifying plants, these parts are rarely preserved. Additional sources of plant identifications include graphic depictions of plants on pottery, seals, murals, etc., and the impressions of seeds, husks, or other plant parts on pottery.

Since seeds may also occur in archaeological samples accidentally, or as a result of later contamination, archaeobotanical investigations make use of a number of quantitative measures to assess the probable uses of plants at particular sites. Weber (1991: 58–9) mentions four such measures: density (e.g. the number of seeds of taxon X per liter of soil sampled), percentage or proportion (e.g. percentage of seeds of taxon X among all seeds, or all botanical remains, in a sample or set of samples), purity (proportion of seeds of taxon X in a single sample), and ubiquity ("percentage of samples from a given assemblage which contains a specific taxon", 1991: 59). Used in combination, sometimes with other types of data, these measures can support inferences about the probable use of particular plants at a site. Thus for example, Weber notes that seeds of *Panicum miliare* (little millet), though found in relatively low densities in Rojdi A, ranked high in measures of ubiquity. Combined with data on the modern uses of this plant and its occurrence in wild form in the vicinity of the site, he concludes that "If not cultivation, then surely intensive collection of wild stands was occurring, for the counts and density of *P. miliare* are too high to be explained in any other manner" (1991: 108).

Unfortunately, these sophisticated methods have been applied to archaeological sites in South Asia only since the 1960s, and even during that period not uniformly. Thus there is much less detailed information available for sites excavated earlier, as well as some later sites. Weber (1991: 22) notes that plant remains have been recovered from only six South Asian sites dating before 2600 BCE (pre-Harappan), from 22 sites dating between 2600 and 2000 BCE (the Mature Harappan period), and from 50 sites between 2000 and 1000 BCE. Furthermore, few sites contain more than 10 recognizable taxa – as compared, for example, with Rojdi, which contains 62.

The detailed discussion of plant prehistory in **7.1** contains numerous references to the Harappan site of Rojdi in Gujarat and to Steven Weber's *Plants and Harappan Subsistence* (Weber 1991). This source has been invaluable for the preparation of this chapter, since it is an up-to-date discussion of palaeobotany which uses the latest available methods.[2] The dates assumed for the various periods of occupation at Rojdi are as listed here: period A (2500–2200 BCE), period B (2200–2000 BCE), period C (2000–1700 BCE).

7.1. Historical and etymological background of South Asian crops and crop names

The following sections discuss particular crop plants, providing brief information about their histories of domestication and the probable sources of their South Asian names. The purpose of this section is to bring together botanical, agricultural, archaeological, and etymological information to aid in tracing the prehistory of these plant species.

The discussion is divided into four sections: Early Harappan or Pre-Harappan (**7.11**), Mature Harappan (**7.12**), and Late Harappan or Post-Harappan (**7.13**), which present data on crop plants which have been identified archaeologically; Section **7.14** discusses a number of plant names which are attested or reconstructed in early periods (mostly dateable to the second millennium BCE or earlier), but for which archaeological data are lacking. In each section the plants are listed in alphabetical order under the English name, with the exception of some late additions: see items C11, C12, D23.

Under each item, the first section presents prehistoric, historical, and general information about the plant in question, while the second section provides etymological data. For Indo-Aryan words, the approximate historical depth is indicated by the following notations given in parentheses after the cited OIA form: RV (= Rigvedic), late V (= later Vedic), Epic (= Epic Sanskrit, i.e. *Mahabharata* and *Ramayana*), Class. (= Classical Sanskrit), late Skt (= post-Classical), lex. (= found only in lexicographical sources); see **2.24** for Indo-Aryan chronology. Historical depth for Proto-Dravidian (PD) is estimated to be 2500–2000 BCE, but could be older; Proto-South Dravidian is estimated to be between 2000 and 1500 BCE, and PSD1 between 1500 and 1000 BCE (see **2.4**). Proto-Munda reconstructions are estimated at about 1500 BCE.[3] The (most probable) source of each plant name is given where information is available; if no source is mentioned, the origin is unknown. Proto-Dravidian reconstructed forms in brackets, such as **[*_cīnt(t)_-]** or **[*_var-inc_-]**, are from Krishnamurti (2003: §1.2); other reconstructed Dravidian forms are the author's.[4]

7.11. Pre-Harappan/Early Harappan (before 2500 BCE): items A1–A6

A1. Barley (Hordeum vulgare)[5]

Historical/general. Barley was domesticated from wild races found today in southwestern Asia. It was one of the earliest crops domesticated in the Near East, perhaps as early as 8000 BCE on the Euphrates. In the city-states of Mesopotamia and in dynastic Egypt, barley remains were much more abundant than those of wheat, and the greater importance of barley is confirmed inscriptionally for both areas. Barley spread eastward to the Indus and to China (J. R. Harlan in Simmonds 1976: 93–5).

In South Asia, impressions of two-row hulled barley (*Hordeum distichum*) and six-row barley (*H. vulgare* and *H. vulgare var. nudum*) were found at Mehrgarh in Baluchistan in contexts dateable to the end of the sixth millennium BCE (Jarrige 1982). Barley was an important crop in the Mature Harappan period at sites like Harappa, Mohenjodaro, and Kalibangan (Allchin and Allchin 1982: 191). Records of barley from the Gangetic plain (Atranjikhera in Uttar Pradesh and Chirand in Bihar) are more or less contemporary with these Indus Valley finds (Randhawa 1980: 163). Weber also notes the presence of small quantities of barley seeds at Rojdi, and concludes that since the amounts are too small to suggest that barley was a significant feature of the diet, its presence in this context probably indicates "direct or indirect contact with people further to the north" (Weber 1991: 106).

Etymological

(A) OIA **yava** (RV) 'barley' (source: IE);[6]

(B) PSD1 ***koc-/*kac-** (Ko **kaj**, To **koj**, Pkt **gajja** DEDR 1106.

A2. Cotton (Gossypium spp.)

Historical/general. The short-staple cottons, *G. herbaceum* and *G. arboreum*, have a long history of cultivation in the Old World. The oldest specimen of cotton cloth is from the Indus Valley, dateable to about 5000 BP. *G. herbaceum* was probably first cultivated in the Arabian peninsula and Syria, whence it found its way to India and gave rise to *G. arboreum*. World production of cotton today is made up mostly of New-World long-staple cottons, mainly *G. hirsutum* (nearly 95 percent of modern production, mostly from USA and the former USSR), with extra-long staple *G. barbadense* accounting for about 5 percent, and short-staple cotton from South Asia accounting for less than 1 percent (L. L. Phillips in Simmonds 1976: 196–8).

The first evidence of the cultivation of cotton comes from Mehrgarh (Baluchistan) in the sixth millennium BCE (Allchin and Allchin 1982: 109), more than a millennium before the date of the piece of dyed cotton (a variety of *G. arboreum*) found at Mohenjodaro (Randhawa 1980: 179). The Greek name for cotton (**sindōn**) points to the Indus Valley as the source. Randhawa claims that cotton cloth was "the Harappans' main export to Mesopotamia" (1980: 180). Ratnagar, on the other hand, states that "There is well-established evidence for cotton spinning and weaving at the Harappan sites, but there is no indication in the Mesopotamian texts that cotton or cotton cloth came to Ur by sea" (1981: 79). Ratnagar also notes that, although both cotton and linen cloth were known to the Harappans, it cannot be assumed that the boats which carried goods to Mesopotamia had cloth sails, since sails of rush matting were known in the region, even in historical times (1981: 164).

Etymological

(A) OIA **tūla** 'tuft of grass, etc.' (late V), 'cotton' (Epic); cf. PSD ***tū-** 'feather, down'[7] DEDR 3393;

(B) OIA **karpāsa** (late Skt) ‘cotton’. Probable source: Austro-Asiatic;[8]
(C) PSD **[*_par-utti_]** ‘cotton, *G. herbaceum*’ DEDR 3976;
(D) PSD **[*_nūl_]** ‘cotton thread’ DEDR 3726.

A3. Date (Phoenix dactylifera, P. sylvestris, P. acaulis)

Historical/general. The date is believed to have been domesticated simultaneously in various places between the Indus and the Atlantic. The earliest record of cultivated dates outside of South Asia is a sample of stones from Egypt dated in the mid-fifth millennium BCE (Simmonds 1976: 230). In South Asia there is evidence for date cultivation at Mehrgarh in Baluchistan before 5000 BCE (Allchin and Allchin 1982: 108), and in the Indus valley during the Mature Harappan period (Marshall 1931: 27, 587; Vats 1940: 467; Allchin and Allchin 1982: 191). The region of Sindh (southern Pakistan) is especially well known for dates. See further discussion in Ratnagar (1981: 80).

Etymological
(A) PD **[*_cīnt(t)-_]** ‘date’ (tree/fruit) DEDR 2617 – source unknown (but cf. item (G) below and proto-Bantu **mu-kindu**);[9]
(B) PD **[*_uc(c)-utt-_]** ‘date’ (fruit) DEDR 620;
(C) OIA **kharjūra** (late V) ‘date’;
(D) KM ***Vn-ḍeñ** ‘*P. Sylvestris/acaulis*’ (not < Drav. acc. to Zide and Zide 1973: 12);
(E) SM ***raloXg** ‘id.’ (Kh. **larog**, Re. **laruk’**, Ga. **ralo?**) (Zide and Zide 1973: 12);
(F) CM ***ḍag** (Zide and Zide 1973: 12);
(G) NM: Mundari-Santali-Ho **kita**, Mundari **kindad/kendad (-dad = CM *ḍag**?) (Zide and Zide 1973: 12).

A4. Grape (Vitis)

Historical/general. (A) Domestication of the grape is believed to have taken place *c.*4000 BCE in Central Asia, in an area running from northeastern Afghanistan to the southern borders of the Caspian and Black Seas, where the single wild species, *V. vinifera*, is still found (Simmonds 1976: 295). A grape seed was among the finds at Loebanr in Kashmir, at a level dateable to the late third millennium BCE (Allchin and Allchin 1982: 115).

Etymological
(A) OIA **drākṣā** ‘grape’ (Class.), with cognates mainly in northwestern and western Indo-Aryan.

A5. Jujube (Zizyphus jujuba, Z. mauritiana)

Historical/general. The jujube, or **ber**, probably originated in Asia, possibly in South Asia (Weber 1991: 100). Remains of the fruit were found at Mundigak in

Baluchistan in a context dateable to the fourth millennium BCE (Allchin and Allchin 1982: 103), though Vishnu-Mittre and Savithri state that *Z. mauritiana Lam* is the true **ber** (1982: 206). Weber notes the presence of carbonized seeds and fruits of *Z. mauritiana* in a single sample at Rojdi A (1991: 108), and suggests that these were probably from a wild rather than a cultivated source.

Etymological

(A) OIA **badara** 'jujube (fruit)' (late V), with cognates in major NIA languages including Sinhala;
(B) PD **[*ira-tt/nt/nk-]** *'Z. Jujuba'* DEDR 475;
(C) PSD ***koṭ-ṭay** DEDR 2070.

A6. Wheat (Triticum compactum, T. sphaerococcum, T. diococcum, T. durum/aestivum)

Historical/general. "The earliest known grains of domesticated wheats date to approximately 7500–6500 [BCE]...in sites of the 'fertile crescent' (Moshe Feldman, in Simmonds 1976: 124). In South Asia, grains of fully domesticated wheat have been reported from the early neolithic levels of Mehrgarh, dateable in the late sixth millennium BCE (Jarrige 1979, 1982; Weber 1991: 29), and from Mundigak (Baluchistan) for the fourth millennium BCE (Casal 1961). Wheat was a major crop of the Harappan civilization (Allchin and Allchin 1982: 191).

Etymological

(A) OIA **godhūma** 'wheat' (late V);[10]
(B) PD **[*kūl-i]** 'rice/wheat' DEDR 1906; see **3.23(1)**.

7.12. The Mature Harappan period (2500–2000 BCE): items B1–B9

B1. Finger millet (Eleusine coracana, E. indica africana, E. indica indica)

Historical/general. While the relationships among the species and subspecies of *Eleusine* are still under discussion, it is believed that the cultivated varieties, including the cereal *E. coracana* (known as **ragi** in India), are all derived from *E. indica africana*, which is confined to Africa (mainly the uplands of eastern and southern Africa). The center of origin may have been in the vicinity of Uganda (J. W. Pursglove in Simmonds 1976: 91ff.).

"*Eleusine coracana* is found in the earliest levels of [the Harappan site of] Rojdi, implying its use as early as 2600 BC in Gujarat" (Weber 1991: 155). Carbonized ragi seeds made up over 62 percent of all the botanical material in phase A at Rojdi. These seeds were collected from all areas of the site, and made up close to 90 percent of the seeds in some samples, indicating that it was one of their major food crops. Ragi is notoriously hardy, providing a reliable staple crop even in drought years, and can be stored for decades without deterioration from pests or climatic

conditions (Weber 1991: 106–7). Carbonized seeds of ragi were also reported from the neolithic site of Hallur, in Karnataka, dateable to about 1800 BCE (Vishnu-Mittre 1974). See **8.32, 8.4** for further discussion of ragi in South India.

Ragi is a major crop in southern (and to some extent western) India, with the states of Maharashtra, Karnataka, Andhra Pradesh, and Tamilnadu accounting for more than two-thirds of Indian ragi production (Randhawa 1980: 235–6).

Etymological

(A) OIA **maḍaka** '*E. coracana*' (lex.) CDIAL 9728;[11]
(B) PSD ***ira(k)** DEDR 812 '*E. coracana*' DEDR 490;
(C) KM ***ḍeray** (also means 'grain in general' in Gorum) (Zide and Zide 1973: 8–9).[12]

B2. Indigo/anil (Indigofera spp.)

Historical/general. *I. tinctoria*, from India and Srilanka, is the principal source of commercial indigo, while other important species are *I. anil* (southern Asia), *I. arrecta* (Africa), and *I. sumatrana* (Southeast Asia). Indigo has now been largely replaced by synthetic aniline (P. M. Smith in Simmonds 1976: 312).

Seeds of at least four different species of the genus *Indigofera* were recovered from the site of Rojdi. One group of mostly carbonized seeds, which could not be identified as to species, was found at levels A and B. Over 50 species of this genus grow in India, and about half of these are found in Gujarat. Some species are considered weeds, while others are cultivated for indigo dye or as fodder, famine food, green manure, or for medicinal uses (Weber 1991: 78–9).

Etymological

(A) OIA **nīla** (RV) 'blue, indigo plant' (with cognates in all major NIA languages including Sinhala and Maldivian), probably connected with PSD1 ***aṇil-** 'ink-nut tree' DEDR 119;
(B) PSD ***aviri** '*I. tinctoria*' DEDR 269;
(C) PSD1 ***koẓinci** 'purple wild indigo' DEDR 2145.

B3. Lamb's quarters (goosefoot, pigweed) (Chenopodium album)

Historical/general. This plant is an annual herb which grows wild and as a weed in cultivated fields in Gujarat and elsewhere in India. It is a prolific seed producer. The young shoots of the plant are eaten as greens, and the seeds are often eaten as a cereal, sometimes mixed with other grains. The seeds may also be mixed with other grains such as ragi or maize to make a fermented drink (Weber 1991: 67–8).

C. album is an important seed in the plant record of the site of Rojdi (periods A and B), where according to Weber's quantitative analysis it has a profile similar to those of the millets *Eleusine coracana* (ragi, item B1), *Panicum miliare* (little millet, item B5), and *Setaria italica* (foxtail millet, item C6) (Weber 1991: 142ff.).

Etymological

(A) OIA **vāstuka** ‘of the house-site’ (= ‘yard-weed’?) (late Skt; Hindi–Urdu **bathuā**, Masica 1979: 86);
(B) PSD ***kīrai** ‘greens, *Chenopodium/Amaranthus*’ DEDR 1617;
(C) PSD1 ***vēḷai** ‘*C. album*, potherb, sticky plant’ DEDR 5546.

B4. Lentil (Lens culinaris)

Historical/general. The wild progenitor of the cultivated lentil is *L. orientalis*, found mainly in Turkey, Syria, Israel, northern Iraq, and northern and western Iran, which is also the area where the first archaeological evidence of lentil cultivation has been found. Lentils are associated with the Old World ‘agricultural revolution’, and carbonized seeds have been found in early near-eastern farming villages dated in the sixth and seventh centuries BCE. Lentils accompanied the neolithic and bronze-age expansion of wheat and barley cultivation in the near east, the Mediterranean basin, and central Europe. From the bronze age on, lentils remained associated with wheats in the area of Mediterranean type agriculture (from north-western India to Spain and North Africa). The main centers of genetic diversity are Afghanistan, Turkey, and Ethiopia (D. Zohary in Simmonds 1976: 163–4).

Lentils were among the many samples of seeds identified at early Indus Valley sites such as Rahman Dheri, dateable to the mid-third millennium BCE. Lentils were also found at Loebanr III in Kashmir, along with barley, wheat, and rice; these finds have been radiocarbon dated in the early centuries of the second millennium BCE (Allchin and Allchin 1982: 115). The first finds in peninsular India are at Navdatoli in period II (mid-second millennium BCE), accompanied by other pulses (Allchin and Allchin 1982: 267). Randhawa notes that the lentil (OIA **masūra**) is excluded from the ingredients of food prepared for the **śrāddha** ceremony, an ancient Hindu rite honoring the dead, and that it was not acceptable as a sacred gift to a Brahman until after 1000 CE, both facts suggesting that it was believed to be of foreign origin. It was also used as fodder for horses after 800 CE (Randhawa 1980: 243–4).

Etymological

(A) OIA **masūra-** ‘lentil’ (late V).[13]

B5. Little millet (Panicum miliare), *common/hog millet* (P. miliare)

Historical/general. *P. miliaceum* is an ancient crop; though cultivated at one time in Europe, by the early Lake Dwellers and others, and known to the early Romans, it is now grown mainly in eastern and southern Asia. Central or eastern Asia, as well as India and the eastern Mediterranean, have been suggested as centers of origin. It has very low water requirements and can be grown in a wide range of soils (P. M. Smith, in Simmonds 1976: 308–9). The plant grows wild in north India (Weber 1991: 85).

Weber states that seeds of *P. miliare* made up the second largest percentage of seeds recovered at Rojdi A (after *Eleusine coracana*, item B1), and were observed in more samples than any other plant. Since the plant grows wild in the region, it is uncertain whether or not these seeds represent a cultivated crop, though it appears likely that the plant was either cultivated or intensively collected by the population of Rojdi, primarily for human consumption and secondarily for animal fodder. Weber emphasizes the importance of this conclusion, since it indicates that "non-African millets were being used in South Asia some 4500 years ago" (Weber 1991: 107–8).

Etymological

(A) OIA **cīna(ka)** 'millet, *P. miliaceum*' (lex.);[14]
(B) PSD ***var-ak-** 'common millet; *Paspalum frumentaceum/crus-galli; Panicum miliaceum*'[15] DEDR 5260; see also **C8**;
(C) PM ***ə-rig** (Zide and Zide 1973: 8).

Millets. The term 'millet' is used for a group of botanically unrelated grasses which share certain features: they generally have coarse grains and are hardier than the major cereal crops. The following items come under this heading: finger millet (*Eleusine coracana*, item B1), barnyard millet (*Echinochloa frumentaceum*, *E. colonum*, item D18), sorghum (*S. bicolor*, item C10), kodra millet (*Paspalum scrobiculatum*, item C8), foxtail millet (*Setaria italica*, item C6), bulrush millet/pearl millet (*Pennisetum typhoides*, item C1), little millet (*Panicum miliare*, item B5), and common millet or hog millet (*Panicum miliaceum*, item B5). See Weber 1991: 83–4, 171–2 for further information.

B6. Mustard (Brassica spp.)

Historical/general. *B. juncea*, the most common cultivated variety in South Asia, is believed to have originated in the Central Asia–Himalayan area; it was known as 'Indian mustard' in medieval Europe. It has long been bred as an oilseed in South Asia (J. S. Hemingway in Simmonds 1976: 57–8). The main area of cultivation of mustard in India today is in the northwest (Panjab, Haryana, Rajasthan, and western Uttar Pradesh).

Seeds of *B. juncea* have been recovered from Chanhu-Daro, a Harappan site in Sind (Randhawa 1980: 179; see also Allchin and Allchin 1982: 191). Weber notes the recovery of two seeds belonging to the genus *Brassica* at Rojdi, one of which has been tentatively identified as *B. campestris* (1991: 66).

Etymological

(A) OIA **sarṣapa** 'mustard' (late V), probably linked to item (C);
(B) OIA **rājikā** 'mustard' (late Skt);
(C) PSD ***ay-a- ←*caca-** DEDR 921 (see 'Mustard' and the accompanying note in Chapter 8, Appendix B, Section B).

B7. Peas (Pisum spp.)

Historical/general. Though neither the wild progenitor nor the early history of the pea is known, carbonized pea seeds have been found in near eastern and European neolithic sites as early as 7000 BCE (with smooth surfaces, indicating that the cultivation of the pea is as old as that of wheat and barley). Vavilov (1949) considered Ethiopia, the Mediterranean, and Central Asia to be likely centers of origin. Peas are grown mainly in the cooler regions of the world, though they have some role as a winter crop in hotter regions (D. Roy Davies, pp. 49ff. in Simmonds 1976). Seeds of field peas (*Pisum arvense*, *P. sativum var. arvense*) were found at Harappa (Vats 1940; Randhawa 1980: 166). Weber notes that six seeds found in level C of Rojdi have been placed tentatively in the genus Pisum (1991: 86–7, 155).

Etymological. Proto-NIA ***maṭṭara** is assumed in CDIAL to account for NIA forms like Hindi **maṭar** 'pea' (found in northwestern and eastern NIA), though there is no early attestation. Gujarati and Marathi have **maṭh**, suggesting a possible link with words like Hindi **moṭh** 'a bean or vetch, Phaseolus aconitifolius' ← OIA **mukuṣṭha**, attested in Buddhist Hybrid Sanskrit (CDIAL 9724, 10148).

B8. Rice (Oryza sativa)

Oryza sativa (with variants), *O. coarctata*, *O. oficinalis*, *O. perennis*, *O. granulata* etc.; wild varieties include *O. fatua* and *O. rufipogon*, among others.

Historical/general. T. T. Chang (in Simmonds 1976: 98ff.) locates the "primary centre of domestication" of rice in a 2,000-mile-long belt which extends from the Ganga–Yamuna doab through northern Bangladesh, and across northern Burma to the Mekong region. Note that though the relationship between *O. sativa* or Asian rice and *O. glaberrima* or African rice is unresolved, it must be assumed that the two cultigens represent two "independent and parallel domestications" (T. T. Chang in Simmonds 1976: 99).

In South Asia, remains of rice are dateable to the fifth or sixth millennium BCE at Koldihwa in Uttar Pradesh, though this early date is uncertain.[16] The next earliest date is in Lothal, a Harappan site in Gujarat, at about 2300 BCE, followed by Rangpur at about 2000 BCE (Randhawa 1980: 272). Later finds include the early second millennium BCE at Loebanr in Kashmir and Chirand in Bihar, somewhat later in Ahar in Rajasthan (Randhawa 1980: 151), in the early second millennium at Chirand in Bihar (Allchin and Allchin 1982: 318–19), and in the mid-second millennium BCE in several sites in Bengal (op. cit. 258, 261). Randhawa states that possible centers of origin in India include the Malabar coast in Kerala, the Jeypore tract in Orissa, and the northeastern foothills of the Himalayas (1980: 273). Weber comments: "Rice is most commonly associated with second millennium sites and in regions where summer [June–October] cultivation is practiced today" (1991: 26).

The early finds of rice in Gujarat were in the form of impressions of spikelets on pottery. As noted earlier, this evidence does not prove that rice was cultivated, or even gathered for food, in these settlements. Weber considers the absence of rice at Rojdi to be "intriguing", since it is grown in the region nowadays and has been found in sites in Gujarat and elsewhere which are contemporary with Rojdi (1991: 156). Randhawa points out that Gujarat, a dry area, is not suited for rice-growing, and believes that the finds in eastern India may be of more significance in tracing the history of this crop (1980: 271). P. G. Chatterjee (1985) states that "primitive rice cultivation" is carried out in two regions in modern India, one in eastern Andhra Pradesh and one in northeastern India (Khasi–Garo area). In addition to these two areas, swidden cultivation of rice was practiced in parts of northern Kerala into the 1960s or later.[17]

Etymological

(A) OIA **vrīhi** 'rice' (late V). See item (C) below and note.
(B) MIA **caula/cavala**; (?connected with PD ***caval** 'mortar' DEDR 2139);
(C) PD **[*var-inc]** – see **3.23(3)**[18];
(D) PD ***key-c-** 'paddy' (unhusked rice) DEDR 1936;
(E) PD ***maṇṭ-** 'rice' (cooked) DEDR 4679;
(F) SM ***ruŋ-kug** (with AA cognates, Zide and Zide 1973: 7);
(G) NM ***baba** (with Mon-Khmer cognates, Zide and Zide 1973: 8).

B9. Sesame (Sesamum indicum)

Sesamum indicum is the principal cultivated species.

Historical/general. Though there are numerous archaeological and ancient literary references to sesame from Mesopotamia, Iran, China, and Malaya, none of these places seem likely candidates for the earliest domestication of this crop. It was long assumed that the original source was located in tropical Africa, because the majority of *Sesamum* species are found there and also because of the importance of sesame in the local economies. N. M. Nayar states,

> sesame could have originated in either the Ethiopian region or in peninsular India or even in both independently. The two regions could be considered botanically contiguous as they share several species and genera, have long had cultural and commercial contacts over both land and sea, and have at least two wild species . . . in common. In addition, a wild taxon . . . which is completely interfertile with the cultivated indicum, occurs on the Malabar coast.
>
> (quoted with permission from N. M. Nayar, 'Sesame', in N. W. Simmonds (ed.), *Evolution of Crop Plants*, London: Longman, 1976, p. 232)

Sesame seeds were found at Harappa, one of the major sites of the Indus Valley civilization (Vats 1940). Sesame was important both as a food and as a component of religious ceremonies in ancient India, and is still so today. Ratnagar notes that sesame was cultivated by the Harappans (1981: 52, note 30) and that sesame oil was used in the Ur III period as a commodity in trade between Mesopotamia and Meluhha (1981: 80), which she and a number of other scholars have identified with the Indus Valley (see Thapar 1975).

Etymological

(A) OIA **tila** 'sesame' (late V); origin unknown;[19]
(B) SD1 ***eḷḷu** (DEDR 854) ← Akkadian **ellu** (Bedigian and Harlan 1986);
(C) PD **[*nū(v)-]** DEDR 3720;
(D) SM ***miñ/*meŋ** (Zide and Zide 1973: 14).

7.13. Late Harappan/Post-Harappan (2000–1700 BCE): items C1–C13

C1. Bulrush millet/pearl millet (Pennisetum typhoides)

This is known in north India as bajra (Hindi **bājrā**).

Historical/general. This crop probably originated in western tropical Africa, the greatest area of variability, whence it was taken, probably by 1000 BCE, to India where a second center of variability developed (J. W. Pursglove in Simmonds 1976: 91–2). In India, bajra is grown mainly in the drier areas of the west and south (Rajasthan, Gujarat, Maharashtra, Andhra Pradesh, and Tamilnadu).

Early identifications of *P. typhoides* in South Asia are unclear. Allchin and Allchin refer to "some millet, possibly *bājrā*" at Rangpur III in the first half of the second millennium BCE (1982: 245), and "a possibility that *bājrā*...was cultivated" at Ahar in Rajasthan in the mid-second millennium (1982: 264). Finds of pearl millet at Hallur, Rangpur, and Ahar lead one to wonder why the species failed to show up at Rojdi (Weber 1991: 156).

Etymological

(A) Proto-NIA ***bājjara** 'millet', with cognates in all major NIA languages except Sinhala;
(B) PSD ***kam-pu** 'bulrush millet' DEDR 1242;
(C) PM ***gaŋ(-)gay** 'sorghum/bajra/large cereal plant' (Zide and Zide 1973: 8).

C2. Chicklingvetch or grass pea (Lathyrus sativus)

Historical/general. This annual herb is believed to be native to western Asia and southern Europe. It is grown in India today, where it is used as a pulse for human consumption and for fodder, though long-term consumption of it causes lathyrism, a paralytic disease, in both animals and humans. A small number of seeds

were found in the upper levels of Rojdi, which along with linseed and other pulses led Weber to suggest that there may have been an increasing emphasis on cultivation of winter crops in post-Harappan Rojdi (Weber 1991: 79–80, 155–6).

Etymological

(A) Proto-NIA ***khesārī** 'Lathyrus sativus, lentil' T3925 (with cognates mainly in eastern NIA).

C3. Chickpea (Cicer arietinum)

Historical/general. The center of diversity of the genus *Cicer* lies in western Asia (Caucasus or Asia Minor). The earliest identification of chickpea was from a site in Turkey, dated to 5450 BCE. Ramanujam suggests that the progenitor of *C. arietinum* was spread by "Aryans" both westward (toward the Mediterranean) and eastwards overland to India, but no evidence is provided for this claim. He suggests that the lack of a common name in Indo-Aryan and Dravidian may indicate that south India received this plant by sea. The earliest find in South Asia was at Atranji Khera in Uttar Pradesh, dated around 2000 BCE (S. Ramanujam in Simmonds 1976: 158–9; v. also Vishnu-Mittre 1974).

Etymological

(A) OIA **caṇa** 'chickpea' (Epic), with reflexes in Pali, Prakrit, and most of the major NIA languages except Sinhala (origin unknown);[20]
(B) PSD1 ***kaṭalai** DEDR 1120.

C4. Emblic myrobalan or Indian gooseberry (Phyllanthus emblica)

Historical/general. Remains of *P. emblica* were found at Navdatoli in contexts dateable to the early second millennium BCE (Allchin and Allchin 1982: 267). The fruit, which has an extremely sour taste, was traditionally used in tanning and for medicinal purposes.

Etymological

(A) OIA **āmalaka** '*P. emblica*' (late V), with reflexes in Pali and Prakrit, and throughout NIA, including Sinhala;
(B) PD **[*uc-Vr-kkāy]** DEDR 574;
(C) PD **[*nel-]** DEDR 3755.

C5. Flax, linseed (Linum usitatissimum)

Historical/general. Flax appears to have originated from a center in southwestern Asia (India, Afghanistan, Turkestan) from where it spread to the north and west. Vavilov found two primary areas, the Asian one containing mostly oil flax types,

and a Mediterranean one with mostly fiber flax types, and a transitional area covering Asia Minor, the Caucasus, and the Black and Caspian Seas. Flax was associated with the deposits of the Swiss Lake Dwellers, and is known from Egypt as early as 1000 BCE. In South Asia, linseed oil was prescribed in Vedic rituals (A. Durrant in Simmonds 1976: 55–6).

The earliest finds of flaxen thread and linseed in South Asia were in chalcolithic sites of Maharashtra dateable to the late second millennium BCE (Allchin and Allchin 1982: 276). Randhawa notes that the crop is now grown in an area surrounding the sites of the earliest finds, namely Maharashtra, Madhya Pradesh, and Uttar Pradesh (Randhawa 1980: 245). Three carbonized seeds, belonging to the genus *Linum* but so far unidentified as to species, were found in Rojdi C levels. Although two species of *Linum* are found in Gujarat, they are both rare in Saurashtra (Weber 1991: 81; see also 155–6).

Etymological

(A) OIA **atasī** 'flax' (late Skt), ?← **atasa** 'bush' (RV)[21] (origin unknown, but note the resemblance to the following form);

(B) PSD1 ***ak-V-ce*** 'flax' DEDR 3.

C6. Foxtail millet (Setaria italica)

Historical/general. This is important as a grain crop in southeastern Europe, North Africa, and Asia (esp. India and Japan). China seems to be the most likely center of domestication. Weber reports minor finds of seeds of *Setaria* species at Rojdi A, and points out that these species grow well in the area, are sometimes cultivated, and are commonly found as weeds in cultivated fields. Though the number of seeds found at Rojdi A is small, they appear with greater frequency at later stages of occupation, especially at period C where it is the only major cereal and where the locations of finds suggest that it may have been carried in and threshed on site (Weber 1991: 89–92, 131).

Etymological

(A) OIA **kanku(nī)-/kangu(nī)-/ṭangunī** '*S. italica, Panicum italicum*, millet' (late Skt);

(B) Proto-IA ***rahala** 'kind of pulse, chickpea, *P. italicum'*, CDIAL 10667;

(C) PD **[****koṯ***-]** '*S. italica, P. italicum*, etc.' DEDR 2163: this Proto-Dravidian word may have earlier referred to bristly foxtail millet (*Setaria verticillata*), a variety of millet grass found in sites of the Southern Neolithic archaeological complex in the mid-third millennium BCE (see **8.41**);

(D) PSD ***ār-V-k-** DEDR 379, see **C8**;

(E) PSD1 ***tiṉai** DEDR 3265;

(F) PM ***(h)oXy** (Zide and Zide (1973: 8), see the Zides' note on the lack of early evidence for this grain, and cf. the discussion of *Setaria italica* and *Setaria verticillata* in **8.41**).

C7. Horse gram [Macrotyloma uniflorum (Lam.)]

This was formerly *Dolichos biflorus, V. sinensis*).[22]

Historical/general. Remains of D. biflorus were found in Karnataka (Tekkalakota I, Hallur II, Paiyampalli), dateable to about 2100 BCE (Allchin and Allchin 1982: 291). Somewhat later are the finds in the Malwa culture in Maharashtra, which were accompanied by wheat, barley, sorghum, field peas, and lentils (op. cit. 273, 276). Nine carbonized seeds of *V. unguiculata* were found at Rojdi, associated with level C and the Early Historic occupation level. Horse gram is still grown in all these areas today. It is believed to be indigenous to South Asia (Weber 1991: 96).[23]

Etymological. The three forms in A–C below are probably all ultimately from the same (unknown) source:

(A) OIA **kulatthā** 'horsegram' (Epic);
(B) PD **[*****koḷ]** DEDR 2153;
(C) PM ***koḍaXj** (Zide and Zide 1973: 10);
(D) PM ***rVm-** 'green gram, cowpea' (Zide and Zide 1973: 10);
(E) Hindi **lobiyā** 'cowpea, *V. sinensis sinensis*' ← Persian **lūbiya** ← ? (Masica 1979: 80).

See also **(C9) Mung, (C13) Urad.**

C8. Kodra millet (Paspalum scrobiculatum)

Historical/general. This is a hardy and drought-resistant crop, currently grown mainly in upland areas in the Indian states of Tamilnadu, Karnataka, Maharashtra, Gujarat, Madhya Pradesh, and Uttar Pradesh. Remains of *P. scrobiculatum* were found in Rupar (Panjab), dateable to 3000–2500 BCE, and this was the only cereal identified at the site of Nevasa (Maharashtra), at levels dateable to the mid-second millennium BCE (Vishnu-Mittre 1969). According to Weber (1991: 86), it is thought to be native to India. Four carbonized seeds of *P. scrobiculatum* were associated with Rojdi C strata, but may represent recent contamination (1991: 85–6).

Etymological

(A) OIA **kodrava** '*P. scrobiculatum*' (Epic);[24]
(B) PSD ***ār-V-k-** '*Panicum italicum/Paspalum scrobiculatum/Setaria italica*' DEDR 379; see **C6**;
(C) PSD ***var-ak-** 'common millet; *Paspalum frumentaceum/crus-galli; Panicum miliaceum*' DEDR 5260; cf. OIA **varuka** 'an inferior grain'; see also **B5**.

C9. Mung (Vigna radiata)

It is also known as 'green gram' in South Asia. One of the most important pulses grown in South Asia today, mung is believed to have originated either in South or Central Asia. It is often sown with sorghum (Weber 1991: 97). Charred grains of

mung were found at Chirand in Bihar, dateable to the early second millennium BCE (Sankalia 1974: 307), and mung was identified at Paiyampalli in Karnataka and Navdatoli in Maharashtra (Allchin 1979b). Five seeds from the site of Rojdi, associated with the Early Historic occupation, or possibly Rojdi C, were identified as *V. radiata*. All of these seeds, from four different samples, were found in association with sorghum (Weber 1991: 97).

Etymological. The ultimate origin of the following names is unknown:
(A) OIA **mudga** 'green gram' (late V), with reflexes throughout NIA including Sinhala;
(B) PD **[*pac-Vṯ/Vl]** 'green gram' DEDR 3941.

See also **(C7) Horse gram, (C13) Urad**.

C10. Sorghum (Sorghum bicolor, Sorghum vulgare), *known in north India as jowar* (*Hindi–Urdu j(u)vār*, *Marathi jvāri, etc.*)

Historical/general. Cultivated sorghums were developed in Africa, and probably moved to India either overland or by coastal trade (H. Doggett in Simmonds 1976: 112–15). Weber (1991: 93) notes that Arabia, Burma, and India are also considered possible centers of origin by some authorities. Sorghum seeds were found at Rojdi in Gujarat in levels C (dateable to 2000–1700 BCE) and D (early historic period), as noted in Weber (1991: 93–4). Sorghum is also reported from Ahar in Rajasthan in the early second millennium BCE, from neolithic settlements in Maharashtra at about the same period, and from Pirak in Baluchistan in the mid-second millennium (Allchin and Allchin 1982: 234, 264, 273, 278). Jowar, one of the most important food grains in India, is grown primarily in Maharashtra and Karnataka at the present time. Weber considers the discovery of Sorghum in this post-Harappan context to be one of the most significant finds at Rojdi, as it is the earliest occurrence of this African crop plant in South Asia, and may have implications for changes in subsistence strategies at this period (1991: 133, 155).

Etymological
(A) OIA **yavākāra** 'barley-shaped' (late Skt) has reflexes in most major NIA languages except Sinhala;
(B) OIA **yavanāla** '*Andropogon bicolor*' (late Skt) has reflexes in all NIA regions except Sinhala (origin: see item (C));
(C) PD **[*co<u>nn</u>a-l]** 'maize, great millet, *S. vulgare* DEDR 2896 is probably derived from an earlier ***coḷ-nel** (cf. PSD ***nel** 'rice in the husk' DEDR 3753); the origin of the first element ***coḷ** is unknown. OIA **yavanāla** is probably a Sanskritized version of PD ***co<u>nn</u>al**, influenced by OIA **yava** 'barley', item A1. This Proto-Dravidian word may have originally referred to browntop millet (*Brachiaria ramosa*), a staple cereal of the Southern Neolithic archaeological complex of the mid-third millennium BCE (see **8.41**);
(D) PM ***gaŋ(-)gay** 'sorghum/bajra/large cereal plant' (Zide and Zide 1973: 8).

C11. Fig (Ficus spp.)

Historical/general. The center of origin of *F. carica*, the principal cultivated variety, is the fertile region of southern Arabia, but the date of domestication is unknown (W. B. Storey, in Simmonds 1976: 205). The Indian pipal tree, *F. religiosa*, which produces an edible but commercially unexploited fruit, is depicted frequently on seals of the Indus Valley culture, and clearly had socio-religious significance; in fact, Parpola notes that all the principal varieties of fig are worshiped as sacred in India (Parpola 1994: 258; see also index, p. 366, under 'pipal'). Weber notes that seeds belonging to the genus *Ficus* were recovered from Rojdi, but have not been identified as to species; of over 65 different species known to exist in India, about one-third can be found in Gujarat (Weber 1991: 76).

Etymological

(A) OIA **pippala** 'berry, esp. of *F. religiosa*' (RV), **pippali** 'berry' (late V), '*Piper longum*' (Epic) (see D15);
(B) OIA **aśvattha** '*F. religiosa*' (late V, with reflexes in eastern NIA and Sinhala); possibly connected with item (F);
(C) OIA **udumbara-** '*F. glomerata*' (late V, with reflexes in all major NIA languages) (← Dravidian, see **3.22A3**, also item (H));
(D) OIA **nyagrodha** 'banyan, *F. indica*' (late V, reflexes in Pali, Prakrit, Ashokan edicts, and Sinhala);
(E) OIA **vaṭa** 'id.' (Epic, all NIA except Sinhala);
(F) PSD ***att-i** (← ***arti**?) '*F. glomerata, F. religiosa*' DEDR 144;
(G) PD **[*cuv-]** '*F. infectoria, F. religiosa*' DEDR 2697;
(H) PD **[*uc(c)-utt-]** 'fruit of the date, *Phoenix dactylifera*' DEDR 620;
(I) PD **[*mēṭi]** '*F. glomerata/racemosa*' DEDR 5090;
(J) PD **[*tōnk-]** '*F. glomerata*' DEDR 3537;
(K) PD **[*ar-ac-/-aḷ]** *'F. religiosa*' DEDR 0202;
(L) PD **[*āl-]** '*F. bengalensis*'(?) DEDR 0382;
(M) PM ***lawa** 'wild fig' (prob. *F. glomerata Roxb.*) (Zide and Zide 1973: 7).

C12. Alfalfa, lucerne (Medicago sativa)

Historical/general. Alfalfa in its primitive state is found in Iran, eastern Anatolia, and around the Caspian Sea. The origin of its cultivated form is associated with the growing importance of horses, probably during the early second millennium BCE, in western Iran. It is mentioned, under the name **aspasti** (← OPers. **aspo-asti** 'horse fodder'?), in a Babylonian text of the seventh century BCE. The name *Medicago* is from the Greeks' term for it, which refers to Media, the land of the Medes. It is recorded that the Emperor of China sent an expedition to Iran in 126 BCE to secure specimens of Iranian horses, and also brought back alfalfa seed (K. Lesins, in Simmonds 1976: 165–6). Weber reports a find of 20 carbonized seeds of the genus *Medicago*, of which five were identified as

M. sativa. He notes that this legume is grown throughout the year in Gujarat as fodder for horses (Weber 1991: 82).

Etymological

(A) Hindi **garārī** (Bulcke 1981 s.v.); cf. OIA **gaṇḍālī** 'a kind of grass' lex. CDIAL 4003, ***gaṇḍāsi** 'knife for cutting sugarcane or fodder' CDIAL 4004 (Bihari **gā̃ṛās(ā)**, Bhojpuri **gā̃ṛās**, Hindi **gā̃ṛāsī, gā̃ṛāsā**) [It is doubtful whether this represents an old word for 'alfalfa'];

(B) Brahui **cōṇḍ** 'lucerne', probably connected with PD ***co<u>nn</u>a-l** 'millet, maize'; see (C10) Sorghum, and **yavanāla** 'the grain *Andropogon bicolor*' in **3.23(5)**.

C13. Urad, black gram (Vigna angularis)

This was formerly *Phaseolus mungo.*

Historical/general. This plant is indigenous to South Asia (Weber 1991: 98). Black gram was found in chalcolithic levels at Navdatoli (Allchin 1979b), and 18 seeds were recovered from Rojdi, associated with strata A, C, and C/D (Weber 1991: 98).

Etymological

(A) Proto-IA ***uḍidda** is assumed by Turner (CDIAL 1693) to account for a number of MIA and NIA forms including Hindi **urad/uṛad**, Gujarati **aṛad**, Marathi **uḍid** (there are no northwestern, eastern, or Sinhala forms. (See also Masica 1979: 78);[25] from Dravidian (see (B)).

(B) PD **[*uẓ-untu]** 'black gram' DEDR 690.

(C) PD **[*min-]** 'black gram' DEDR 4862.

See also **(C7) Horse gram, (C9) Mung**.

7.14. Other South Asian crop names (items D1–D25)

This section lists crop plants for which there is no clear archaeological evidence, but which for one reason or another are likely to have been used in the prehistoric or early historic periods. Some appear to have originated and/or been first domesticated in South Asia; others have ancient names, and/or are implicated in ancient rituals. In some cases, the plants are such that archaeological identification is unlikely, for example in the cases of sugarcane, yams, and coconut, which are reproduced vegetatively and do not produce seeds. Where the only evidence for ancient cultivation is etymological, inclusion has been restricted to OIA items from the preclassical period and Dravidian items reconstructible to Proto-Dravidian or Proto-South Dravidian. (A few items from PSD1 have been included; see **8.37**.)

D1. Areca palm (Areca catechu)

Historical/general. The areca tree is widely cultivated in India, Myanmar (Burma), Thailand, and the Malay archipelago, and may have originated in central Malaysia. The nut is sliced and chewed with leaves of betel (*Piper betel*, item **D3**), and has therefore come to be called 'betel nut'. Betel chewing is a "habit of great antiquity" (P. M. Smith in Simmonds 1976: 318).

Etymological

(A) OIA **pūga** 'areca tree/nut'[26] (late Skt): origin: item (C), or a third source;
(B) Proto-IA ***suppāra** 'areca nut' CDIAL 13482, of unknown origin, with cognates in the major NIA languages except Sinhala;
(C) PD **[*pānkk-]** DEDR 4048 'areca (tree/nut)', origin unknown; may be the source of item (A), or both may be from a third source;
(D) PSD **[*aṭ-ay-kkāy]** 'areca nut' (***kāy** 'fruit/vegetable') DEDR 88, the source of Portuguese **areca**, etc.

D2. Banana/plantain (Musa acuminata, M. balbisiana)

Historical/general. Wild varieties of *Musa* occur in an area stretching from southern India through Southeast Asia to New Guinea and northern Australia. Taxonomic evidence indicates that the primary center of evolution was the Malay peninsula, possibly including other nearby areas. The cultivated edible varieties belong to *M. acuminata*, *M. balbisiana*, or to various hybridizations of these. No dating is possible, but the early diffusion of the plant probably occurred millennia ago. Bananas entered Africa before European contact, probably from Malaysia via Madagascar rather than from India. At approximately the same time, the plant was transported across the Pacific (N. W. Simmonds in Simmonds 1976: 212–13). According to Randhawa, *M. balbisiana* is a native of eastern India ("from Bihar, up to the Himalayas") and the principal area of genetic diversity is Assam–Thailand (Randhawa 1980: 266–7).

Etymological

(A) OIA **kadala(ka)/kadalī** 'banana' (Epic), is found in most NIA languages including Sinhala (origin: Austro-Asiatic);
(B) PSD1 **[*vāẓ-a]** DEDR 5373.[27]

D3. Betel (Piper betel)

Historical/general. See item D1; for background, see A. C. Zeven in Simmonds 1976: 234.

Etymological

(A) OIA **tāmbūla** 'betel leaf' (late Skt), **tāmbūlika** 'seller of betel' (Epic);[28] origin: Austro-Āsiatic;[29]
(B) PSD1 ***veṯṯ-ilai** 'betel leaf', DEDR 5515 (PSD ***ilai** 'leaf') is the source of the word Port. **betel**, etc.

D4. Cardamom (Elettaria cardamomum)

Historical/general. Native to India, this plant is grown mainly in southern India and Srilanka (P. W. Smith in Simmonds 1976: 324).

Etymological. The ultimate origin of all the following is unknown:
(A) OIA **elā** 'cardamom' (late Skt) may be a borrowing from (C), or from a third source;
(B) Proto-IA ***troṭikā** is found in the northwestern and some eastern NIA languages, and similar forms are found in Tamil and Malayalam DEDR 3298);
(C) PSD **[**ēl*]** DEDR 930.

D5. Coconut (Cocos nucifera)

Historical/general. The coconut palm originated in the islands of southeastern Asia, in an area centered on Borneo and including the lower Malay peninsula, Sumatra, Java, the Philippines, and western New Guinea. From there it spread westward to Africa and the Atlantic coast of South America, and eastward across the Pacific as far as the South American west coast. The chronology is unknown, but the tree is well enough established in South and Central America to have prompted some botanists to suggest this as the place of origin (R. A. Whitehead in Simmonds 1976: 222–3).

Vishnu-Mittre (1974) has identified a possible specimen of wild coconut in Srilanka, dateable to about 5700 BCE. In India, Kerala is the largest producer of coconuts, followed by Tamilnadu and Karnataka, with small amounts grown on the coasts of Maharashtra and Andhra Pradesh (Randhawa 1980: 392). There is textual evidence to show that the coconut was being cultivated on India's east coast by the middle of the first century BCE, reaching the west coast about a century later, though it was rather poorly known in many parts of India before the sixth century CE (Randhawa 1980: 392). Remains of coconuts have not been identified archaeologically in South Asia, but Vats inferred that the fruit was known in Harappa, on the basis of the shape of an earthenware vase found there (Vats 1940, cited in Randhawa 1980: 180).

Etymological. All the following are of unknown origin:
(A) OIA **nārikela/nārikera/nālikera/nālikela** has reflexes in most major NIA languages (including Sinhala but excluding Bangla);[30]
(B) PSD **[**ten-kāy*]** 'coconut fruit, coconut palm' DEDR 3408;
(C) PSD***koppa-ray** 'copra, coconut kernel' DEDR 2105, cf. Proto-IA ***khoppa(ra)** CDIAL 3936;
(D) PSD***kaiṟu** 'coir, coconut fibre' DEDR 1254.

D6. Cowpea

See C7 Horse gram.

D7. Egglant, brinjal, aubergine (Solanum melongena, S. indicum)

Historical/general. Evidence for early cultivation in India is primarily etymological, though wild *S. melongena* occurs there. The eggplant was known in China in the fifth century BCE. It reached western Europe before the middle ages (B. Choudhury in Simmonds 1976: 278). Kochhar states that the plant appears to have been domesticated in northeastern India (1998:267). Weber reports that 39 of a total of 65 seeds of *Solanum* species found at Rojdi appear to be *S. melongena* (Weber 1991:92).

Etymological. Item (C) has the greatest apparent time depth, but is not necessarily the source of either of the OIA words (both of which have numerous variant forms). Items (A) and (C) were most likely borrowed from a third source.

(A) OIA **vātingaṇa** (lex.) has reflexes in all major NIA languages except Sinhala;
(B) OIA **bhaṇṭākī** (lex.) has reflexes in Pali, eastern NIA, and Sinhala;
(C) PD **[*vaẓ-Vt-]** DEDR 5301; cf. Tamil **vaẓutalai/vaẓutaṇai**, Malto **baṭango**).

D8. Field bean, hyacinth bean (Dolichos lablab)

Historical/general. This crop has long been cultivated in India, and is still grown widely. Wild forms also occur in India, which may be the area of origin. The beans and pods are eaten and the foliage provides fodder and green manure; medicinal uses are also recorded (P. M. Smith in Simmonds 1976: 312).

Etymological

(A) OIA **śaimbya** 'pertaining to legumes' (← **śimba** 'pod, legume'); Hindi **sem** '*D. lablab*';
(B) PSD ***cikk-Vṭ-** DEDR 2496.

D9. Ginger (Zinziber officinalis)

Historical/general. This root crop is probably native to Southeast Asia, though long cultivation has obscured the sites of original domestication and wild distribution. It was used in China and India from ancient times (P. M. Smith in Simmonds 1976: 324).

Etymological

(A) OIA **śṛngavera** 'ginger' (Epic) has reflexes in Pali and Sinhala (probably a classicalization of proto-NIA ***singivera**[31] ← PSD1 ***cinki** (item (C)) + PD ***vēr** 'root' DEDR 5535);
(B) OIA **ārdraka** 'fresh ginger' (late Skt), a descriptive word based on **ārdra** 'wet';
(C) PSD1**[*cink-i]** 'ginger' DEDR 429.

D10. Hemp (Cannabis sativa)

Historical/general. Hemp as a crop originated "somewhere in temperate Asia" (within the large area from the Caspian Sea and the Himalayas to China and Siberia). It is said to have been the only fiber available in parts of ancient China and Siberia, and probably has been cultivated in China for at least 4,500 years. It reached western Asia and Egypt during the second millennium BCE. Its narcotic properties were known in India by 1000 BCE (N. W. Simmonds in Simmonds 1976: 203–4).

Etymological. The following are all from unknown sources, with the possible exception of item (C):

(A) OIA **bhanga** 'hemp' (late V) has reflexes in all the major NIA languages (though note that the meaning in Sinhala and Maldivian is 'intoxicating drink, arrack').
(B) OIA **śaṇa** 'hemp' (late V) has reflexes in Pali and all the major NIA languages including Sinhala (see Witzel 1999b: 55 for possible Central Asian connections).
(C) OIA **ganjā** 'hemp' (lex.) has reflexes in Prakrit and in northwestern and western NIA (CDIAL cites a Sumerian GAN.ZI 'hemp(?)').
(D) PD **[*boy-Vl]** 'hemp fibre' DEDR 4535.
(E) PSD ***kōnk** '*Hibiscus cannabinus*' DEDR 2183 – cf. item (C).
(F) PSD1 ***manc-i** 'hemp' DEDR 4637.

D11. Jackfruit (Artocarpus heterophyllus)

Historical/general. The genus *Artocarpus* is native to Southeast Asia; in the precolonial period, the jackfruit was distributed from Malaysia to India, and its center of origin might be anywhere from southern India to Borneo (Jacques Barrau in Simmonds 1976: 201–2).

Etymological

(A) OIA **kaṇṭaphala** '"thorn-fruit", name of various trees including jackfruit and breadfruit trees' (lex.) has reflexes in eastern and central NIA;
(B) OIA **panasa** 'breadfruit tree' and similar words (Epic), see next item;
(C) PD **[*pal-ac/pan-ac]** 'jack (fruit/tree)' DEDR 3988 may be the source of item (B), or a third source may be involved;
(D) PSD ***coḷ-ay** 'edible part of jackfruit' DEDR 2704;
(E) PSD1 ***cak-ka** 'jack tree' (note PD ***kā(y)** 'fruit/vegetable', DEDR 2275), the presumed source of Portuguese **jaca**, English **jack**, etc.

D12. Lemon, lime, orange (Citrus spp.)

Historical/general. Southern Asia is the place of origin of *Citrus* and related plants, with the main center probably in eastern India. No wild ancestors have

been identified, and the period of domestication cannot be dated. Written records from ancient Iran identify the citron about 300 BCE, while other varieties were not mentioned in writing until after 1000 CE (J. W. Cameron and R. K. Soost in Simmonds 1976: 261–2).

Etymological

(A) OIA **nimbū(ka)** 'lime' (lex.) has reflexes in the main NIA languages (excepting Sinhala) in the meaning 'lime' or 'lemon'; note that the eastern languages have **-e-** in the first syllable (e.g. Bangla **nebu/lebu**). This word shows the fluctuation between initial **l** and **n** which is found in areas of contact with 'tribal' languages (see **5.22D**). The word is probably from an Austro-Asiatic source (see EWA s.v.). Cf. PSD **[*vē-mpu]** 'neem, *Azadirachta indica*' DEDR 5531;

(B) OIA **nāranga** 'orange tree' (late Skt) has reflexes in Pali and in the major NIA languages (again excepting Sinhala); presumably a borrowing from item (C);

(C) PSD1 ***nāram-ka** DEDR 3653 (where the second element is a form of PD ***kā(y)** 'fruit/vegetable');

(D) PSD ***kiccili/kittili** 'orange' DEDR 1512;

(E) PSD ***māt-aḻ** 'citron/lemon' DEDR 4808;

(F) PSD(?) ***īẓ-e** 'orange' DEDR 552;

(G) PSD1 ***elu-mic-cai** 'lemon' DEDR 836 (**elu** = F?).

D13. Mango (Mangifera indica)

Historical/general. The mango probably developed in northeastern India from the wild *M. indica*, perhaps as early as 4000 BP. It moved eastwards with Buddhist monks in the mid-to-late first millennium BCE (L. B. Singh in Simmonds 1976: 7–8).

Etymological. The following words are probably from the same (unknown) source:

(A) OIA **āmra** (late V) has reflexes in all major NIA languages including Sinhala.

(B) OIA **mākanda** 'mango' (Class.).

(C) PD **[*mām-]** 'mango' DEDR 4782; ***mān-kā** '(unripe) mango fruit' is the source of Port. **manga**, E. **mango**, etc.

(D) PD **[*maṯ-kāy]** DEDR 4772.

(E) PM ***uXlax/*uXliX** (Zide and Zide 1973: 7).

(F) PM ***kVyeXr** 'unripe mango' (Zide and Zide 1973: 7).

D14. Okra (Abelmoschus esculentus)

Historical/general. The accepted view is that cultivated okra originated in west Africa, from where it migrated to the Mediterranean region and India, but it is

not impossible that it may have arisen independently in India (A. B. Joshi and M. W. Hardas in Simmonds 1976: 55–6).

Etymological

(A) OIA **bhiṇḍā** (late Skt.) with reflexes in all the major NIA regions except Srilanka.[32]

D15. Pepper (Piper nigrum)

Historical/general. The hills of Malabar, in southwestern India, are presumed to be the center of origin of this crop, because of the presence of wild varieties there, according to A. C. Zeven (Simmonds 1976: 234), though Harlan (1971) regards Southeast Asia as the probable home. The plant has been cultivated in Kerala from about the second century BCE (Randhawa 1980: 406).

Etymological. The similarity between (A) and (B) given here leads to the inference that they are from the same source, probably Austro-Asiatic (cf. Mon **mrak**), which would support Harlan's argument. Items (C) and (D) are also related, with (D) possibly the source of (C), the ultimate origin being unknown.

(A) OIA **marīca** 'peppercorn' (Class.) has reflexes in Pali and all major NIA languages including Sinhala;
(B) PSD1 **[*miḷ-Vku]** 'black pepper' DEDR 4867;
(C) OIA **pippalī** 'long pepper' (Epic, Class.)[33] has reflexes in all major NIA languages (Sinhala has **tipli**, similar to D);
(D) PSD1 ***tipp-a-li** 'long pepper' (← ***tir-pali**?), DEDR 3228.

D16. Pigeon pea (Cajanus cajan, C. bicolor, C. flavus)

Historical/general. Though the origin of this crop has long been under dispute, and there is no direct archaeological evidence, it seems likely that it originated in peninsular India. D. N. De notes that India is the present center of greatest diversity, and he quotes Sanskrit names for the crop from a lexicon and a medical text (the **Suśruta**), both dated in the mid-first millennium CE. The two Sanskrit names, **tubarī** and **āḍhakī**, appear to refer to the two peninsular varieties (*C. flavus* and *C. bicolor* respectively), which seem to be partly differentiated within India, the southern *flavus* varieties having small, quickly maturing plants in comparison with the northern *bicolor* varieties. If Indian origin is accepted, it is probable that the pigeon pea moved from India to Malaysia about 2000 BP. *Cajanus* is derived from the Malaysian name Katjang (W. Vernon Royes in Simmonds 1976: 154–5.)

Etymological. The ultimate source of all the following is unknown:

(A) OIA **tubarī** '*C. indicus'* (lex.), with reflexes in central and southwestern NIA, is clearly connected with item (C), and given the probable age of the latter, the Dravidian word could well be the source of the OIA word;

(B) OIA **āḍhakī** is probably derived from **āḍhaka** 'a measure of grain';
(C) PD **[*tu-var-]** '*C. indicus*' DEDR 3353;
(D) PD **[*kar-Vnti]** 'pigeon pea, tuar pulse' DEDR 1213;
(E) PSD ***ala-cant-*** '*Dolichos catjang/biflorus/sinensis*' DEDR 242.

D17. Pomegranate (Punica granatum)

Historical/general. This plant is native to Iran, and was known to the ancient Romans who associated it with the city of Carthage. It is widely grown and naturalized in the Mediterranean area (P. M. Smith in Simmonds 1976: 308).

Etymological. The following items are of unknown origin:
(A) OIA **dāḍima** (Epic) has reflexes in all major NIA languages including Sinhala;[34]
(B) Tamil–Malayalam **mātaḷam** 'pomegranate' DEDR 4809 (cf. PSD ***māt-aḷ-*** 'citron, lemon' DEDR 4808).

D18. Sawa millet (Echinochloa colonum)

Historical/general. *E. frumentacea* or Japanese barnyard millet is native over a large area of the Old World, from tropical to temperate zones, and is possibly derived from a wild variety native to Java and Malaysia. It is eaten by Hindus on fast days. The related barnyard millet (*E. crus-galli*) is a wild grass which is occasionally cultivated; in India it is used as a forage crop under the name **bharti** (P. M. Smith in Simmonds 1976: 308). *E. colonum* is also known as sawa millet in India. Twelve seeds resembling *E. colonum* were recovered from Rojdi C occupation levels (Weber 1991: 72–3).

Etymological. (A) OIA **śyāmāka** '*Panicum frumentaceum*' (late V) has reflexes in the major NIA languages with the exception of Sinhala.

D19. Sugarcane (Saccharum, S. officinarum)

Historical/general. The New Guinea area, still the primary center of diversity of varieties derived from *S. robustum*, the probable ancestor of the cultivated *S. officinarum* ('noble' sugarcane), is probably the region of domestication. Subsequently the 'noble' canes migrated northwestward to southern China where they hybridized with local *S. spontaneum* to yield the hybrid *S. sinense*, which thrived under the monsoon conditions of northeast India and southern China. No dates, however, can be estimated for these developments (N. W. Simmonds in Simmonds 1976: 104–5). Weber states that many species of *Saccharum* are native to South Asia, and that some species grow wild in Saurashtra, though the burned spikelets and seeds recovered there represent recent contamination (1991: 88). Linguistic evidence from both Indo-Aryan and Dravidian attests the presence of sugarcane around the beginning of the second millennium BCE. Note, however,

that other types of sugar have been and continue to be used in India, such as sugar made from the juice of the palmyra or toddy palm (see item D20) and from the fruit of the jack tree (*Artocarpus heterophyllus*, item D11).

Etymological. All the OIA words for sugar and sugarcane have possible links with Dravidian:

(A) OIA **ikṣu** 'sugarcane', cf. **ikṣvāku** (RV) the name of a Rigvedic clan, (possibly ← Dravidian ***iṭ-cu** ← ***i<u>n</u>-cu** 'sweet-juice': cf. PD ***i<u>n</u>** 'sweet' DEDR 530, and note Telugu **incu** 'sugarcane'), PSD ***cuna** 'juice' DEDR 2717, PD ***cur** 'suck' DEDR 2712;
(B) OIA **gaṇḍa** 'joint of (sugarcane) plant' probably ← PD ***kaṇṭu** 'joint, knot' DEDR 1946;
(C) OIA **śarkarā** 'gravel, grit' (late V), 'candied sugar' (late Skt): Austro-Asiatic origin has been suggested, though a derivation from (D) is also possible, and cf. PSD1 **cakkai** 'jack(fruit)', DEDR 2275; alternatively, the OIA and Dravidian forms might both be from the same (Austro-Asiatic?)[35] source;
(D) PD **[*ce<u>t</u>-]** 'sugarcane' DEDR 2795 (Ka **ceruku**, Kl **saragurak**);
(E) PSD1 **[****kar-umpu***]** 'sugarcane, palmyra sugar' DEDR 1288.

D20. Toddy palm, palmyra (Borassus flabellifer/flabelliformis)

Historical/general. The palmyra grows in India and Queensland; little is known of the origins of this and related species (P. M. Smith in Simmonds 1976: 318). Cultivated throughout India. Though there are no early references, the tree appears frequently in Buddhist sculptures.

Etymological

(A) OIA ***tāḍa**, Pali **tāla** CDIAL 5750 has reflexes in the major NIA languages, including Sinhala and several Gipsy languages (possible source);
(B) PD **[****tāẓ***]** '*B. flabelliformis*' DEDR 3180;
(C) PD ***nunk-u** 'palmyra fruit (tender/unripe)' DEDR 3698.

D21. Turmeric (Curcuma longa)

Historical/general. This plant is native to Southeast Asia, and widely cultivated in India, where wild species also occur. It is used as a spice and dye, and occasionally for food (wild varieties contain considerable starch). It was known in Greece in the first century CE (P. M. Smith in Simmonds 1976: 323).

Etymological. Both NIA and Dravidian have words which are also color terms:

(A) OIA **haridrā** 'turmeric' (late V), with reflexes in major NIA languages including Sinhala;
(B) PSD1 ***manc-aḷ** 'turmeric, yellow' DEDR 4635;
(C) PM ***R-saŋ** (Zide and Zide 1973: 7).

D22. Yam (Dioscorea spp.)

Historical/general. Yams developed independently in Asia, Africa, and tropical America. The major center of diversity of cultivars in Asia is Papua-New Guinea, but northeastern Southeast Asia was probably the earliest area of domestication. Yams were transported across the Pacific in the Polynesian migrations, which started around 1500 BCE. See Fuller *et al*. (2001: 174–5) for the possible presence of some type of yam at various Southern Neolithic sites. Yams are grown throughout the wetter tropics; in India, they are cultivated as a minor crop, mainly in Assam, Bihar, Bengal, Madhya Pradesh, Orissa, and the coastal areas of the peninsula (Kochhar 1998: 237).

Etymological

(A) OIA **raktālu(ka)** '*D. purpurea*' (3,L) (← **rakta-ālu** 'red tuber');
(B) PD ***kiẓ-** 'root, tuber'→ PSD '*Dioscorea aculeata*' DEDR 1578;
(C) PD **[**kic-ampu*]** DEDR 2004.

D23. Bottle gourd (Lagenaria siceraria)

Historical/general. Also known as 'white-flowered gourd', this plant is "most likely indigenous to the tropical lowlands of Africa south of the equator", yet it was present in the western hemisphere as early as 7000 BCE. Remains were found in Egyptian tombs dated to *c*.3500–3300 BCE (T. W. Whitaker and W. P. Benis, in Simmonds 1976: 67).

Etymological

(A) OIA **alābu** (late V) ← Austro-Asiatic;
(B) OIA **tumba** (late Skt.) ← Austro-Asiatic;
(C) PSD ***cur-ai** DEDR 2790;
(D) PM ***su(-)ku(g)**, KM ***N-tuŋ** (Zide and Zide 1973: 11).

D24. Onion (Allium cepa), *garlic* (A. Sativum)

Historical/general. Unknown in the wild state, onions have been cultivated from the earliest times in the region stretching from the eastern Mediterranean to Pakistan. They were cultivated in Egypt at least 4,800 years BP. In India they are grown mainly in Panjab, Bihar, Maharashtra, Tamilnadu, and Andhra Pradesh. Garlic originated in the eastern Mediterranean region. China is currently the world's largest producer of garlic (Kochhar 1998: 240–1). Both onions and garlic are among the foods avoided by some groups of orthodox South Indian Brahmans.

Etymological

(A) PD **[**uḷḷi*]** 'onion/garlic' DEDR 705;
(B) OIA **kanda** 'bulbous root; garlic' CDIA 2723 ('onion' in various NIA lgs) ← PD ***kaṇṭa** 'bulbous root' DEDR 1171;
(C) OIA **laśuna** (also ***raśuna)** 'garlic' CDIAL 10990.

D25. Trees

The following additional names of trees have been included here because they are attested in early Dravidian, either PD (15 items) or PSD (11 items); see Chapter 8 and Appendices A and B.

Attested in Proto-Dravidian (see Chapter 8, Appendix A, Section B):

- Indian horseradish tree, *Moringa pterygosperma*: PD **[**murun-kāy*]** DEDR 4982;
- Tamarind, *Tamarindus indica*: PD **[**cin-tta*]** DEDR 2529; PM ***R-tiXn, *joX/joḍ/joj** (Zide and Zide 1973: 13–14); OIA **ciñcā**, **ciñcini**, **cintiḍī** CDIAL 4792 (← Munda?);
- Palas, *Butea frondosa*: PD **[**mur-ukk-*]** DEDR 4981; OIA **kiṁśuka** CDIAL 3149;
- Cadamba, *Anthocephalus cadamba*: PD **[**kaṭa-(m)p-*]** DEDR 1116 → OIA **kadamba** CDIAL 2710; OIA **nīpa** CDIAL 7549;
- Coomb teak, *Gmelina arborea*: PD ***kūm(p)-iẓ** DEDR 1742 → OIA **kārṣmarya** CDIAL 3082, **gambhārī**;
- Marking-nut, *Semecarpus anacardium*: PD **[**kīr-*]** DEDR 2005; OIA **bhallāta** CDIAL 9414;
- Lac, *Schleichera trijuga*: PD **[**puc-/*puy-*]** DEDR 4348; OIA **alakta** CDIAL 695, **lakṣa** CDIAL 638, **jatu** CDIAL 5093;
- Teak, *Tectona grandis*: PD **[**tēnkk-*]** DEDR 3452; OIA **śāka** CDIAL 12369
- Mahua, *Bassia longifolia/latifolia*: PD **[**ir-upp-a*]** DEDR 0485; OIA **madhūka** CDIAL 9801 (cf. **madhu** 'honey');
- Inknut, *Terminalia chebula* : PD **[**kaṭu-kkāy*]** DEDR 1134; OIA **araṭu** CDIAL 597;
- Laurel, *Terminalia tomentosa*: PD **[**mar-Vt-*]** DEDR 4718; OIA **asana** CDIAL 963;
- Sal, *Shorea robusta*: PD **[**car-*]** DEDR 2420; OIA **śāla** CDIAL 12412;
- Rose-apple, *Eugenia jambolana*: PD **[**ñā(n)ṯ-*]** DEDR 2917; NM ***koXda**, SM ***ku-ḍaX**; OIA **jambu** CDIAL 5131;
- Belleric myrobalan: PD **[*tānṯ-i*]** DEDR 3198; OIA **āmalaka** CDIAL 1247;
- Bamboo, *Bambusa arundinacea*: PD **[**vet-Vr-*]** DEDR 5485; OIA **venu, veḍu** CDIAL 12096, cf. **veta** 'cane, reed' CDIAL 12097 (← Drav?); note the following Munda words: NM ***maXd**, CM ***kaXl**, SM ***kV(-)reXŋ/*kV(-)ruXŋ** (Zide and Zide 1973: 12); OIA **vaṁśa** RV, **vanju** Mbh (← Drav., see **3.22A**); OIA **kamaṭha *kambāṭha kambi *kambiṭṭha**, etc. CDIAL 2760 (← AA); OIA ***conga** 'section/joint of bamboo' CDIAL 4921.

Attested in Proto-South Dravidian (see Chapter 8, Appendix B, Section B):

- Soapnut, *Sapindus emarginatus*: Vahl. **[**cīk-kay*]** DEDR 2607a; OIA **ariṣṭa** CDIAL 610;

- Kino, *Pterocarpus marsupium*: ***vēn-kay** DEDR 5520;
- Agasti, *Agasti grandiflora*: ***aka-c-tti** DEDR 0005 → OIA **agasti** CDIAL 50;
- Pandanus, *P. odoratissimus*: ***kay-tay** DEDR 2026, ***mo-ka-ri/li** DEDR 4890; OIA **ketaka** CDIAL 3462, **jambāla** CDIAL 5129, **jambula** CDIAL 5136
- Clearing-nut, *Strychnos potatorum*: ***cil(l)-** DEDR 2560; OIA **inguda/ingula** CDIAL 1554, **kataka** CDIAL 2691, **tumburu** CDIAL 5872.
- Mastwood, *Polyalthia longifolia* Thw.: **[*puṇṇ-ay*]** DEDR 4343.
- Wood-apple, *Feronia elephantum*: **[*veḷ-*]** DEDR 5509; OIA **kapittha** CDIAL 2749, **jīvalā** CDIAL 5248, **ṭanka** CDIAL 5429, **elavālu** 'bark of *F. Elephantum*' CDIAL 2521;
- Citron lemon, *Citrus medica*: ***māt-aḷ** DEDR 4808 → OIA **mātulunga** CDIAL 10013;
- Neem, *Azadirachta indica*: **[*vē-mpu*]** DEDR 5531; OIA **nimba** CDIAL 7245;
- Oleander, *Nerium odorum*: ***kaṇa-** DEDR 1164; OIA **karavīra** CDIAL 2800; see **3.23(5)**;
- Sandal, *Santalum album*: ***cāntu** DEDR 2448 → OIA **candana** CDIAL 4658.

7.2. Historical implications

The histories of these economic plants and the etymologies of their names provide some clues to communication between different regions and among different sociolinguistic groups. The main conclusions to be drawn from the data given here are discussed next.[36]

7.21. Indo-Aryan and Dravidian

Dravidian languages provided the source of OIA names for the following crops:

Crop name		*OIA form*	*Dravidian source*	*Origin/center of domestication*
(B8)	rice	**vrīhi** (A2)	← PD ***var-inc**	Eastern India–SE Asia
(C10)	sorghum	**yavanāla**	← PD ***co<u>nn</u>-al**	Africa
(D5)	copra	***khoppa(ra)**	← PSD ***koppa-ray**	SE Asia
(D9)	ginger	**śṛngavera**	← PD ***cinki-vēr**	SE Asia
(D12)	orange	**nāranga**	← PSD ***nāram-ka**	Eastern India
(D19)	sugarcane	**ikṣu** (late V)	← PD ***iṭ-cu**	New Guinea
		gaṇḍa (lex.)	← PD ***kaṇṭu**	
		śarkarā	← PD ***ce<u>r</u>-aku**	

The majority of these cases involve plants whose origin lies to the east of India, suggesting that these crops may have passed through Dravidian-speaking areas,

perhaps in coastal regions of the peninsula, before becoming known to speakers of OIA. The same may have been true for sorghum, which could have traveled by sea from east Africa.[37] Of these cases, only rice and sorghum are supported by archaeological evidence; the remaining plants may have entered later, though the linguistic evidence for sugarcane suggests that it was known at least by the late second millennium BCE.

If rice also reached Dravidian-speaking people first by sea, this event may well be independent of the development of rice cultivation in the primary rice belt which extends from eastern India into Southeast Asia (see item B8 in the table). Note that several words for rice are reconstructible for Proto-Dravidian, dateable in the mid-second millennium BCE (Chapter 8, Appendix A, D1). The possibility must be entertained that the rice which was identified at the late Harappan sites of Rangpur and Lothal in Gujarat reached there by sea, either directly or from other coastal parts of India. That Gujarat was once a Dravidian-speaking area is shown by the evidence of place names (Chapter 9).

In another group of cases, OIA and Dravidian have words of similar phonological shape, but it is not possible to determine the direction of borrowing; in some of these cases, the phonological differences are such as to suggest independent borrowing from the same or related sources:

Crop name		*Origin/center of domestication*
(B2)	indigo: OIA **nīla** (RV) : PSD1 ***aṇile***	South Asia?
(C5)	flax: OIA **atasī** (late Skt) : PSD1 ***akace**	Southwest Asia/India
(C7)	horse gram: OIA **kulattha** (Epic) : PD ***koḷ**	India?
(C9)	mung: OIA ***uḍidda** : PD ***uẓ-untu**	India?
(D1)	areca: OIA **pūga** (late Skt) : PD ***pōkku**	Malaysia
(D4)	cardamom: OIA **elā** (late Skt) : PSD ***ēl-**	India
(D7)	eggplant: OIA **vātingana** (lex.) : PD ***vaẓu(ta)**	India/China
(D13)	mango: OIA **āmra** (late V) : PD ***mām**	Northeastern India
(D16)	pepper: OIA **marīca** (late Skt) : PSD1 ***miḷ-Vku** OIA **pippala** (Epic) : PSD1 ***tippali**	Kerala/Southeast Asia

As most of these crops (may have) originated in South Asia, these words probably entered Indo-Aryan and Dravidian – in some cases quite early, for example, eggplant in PD – from one or more indigenous South Asian languages. Munda or other Austro-Asiatic languages are possible candidates for this role (see **7.22**).

In the following case, Old Indo-Aryan appears to be the source of the Dravidian names: (B6) mustard: OIA **sarṣapa** → PSD (***cacavi**.→) ***ay-a . . .** Given that this crop probably originated in central Asia, this is not surprising. (This is one of a number of words found in OIA and Iranian which Witzel (1999b: 55) attributes to an unknown Central Asian source.) Another case of Central Asian origin is the grape (A4) OIA **drākṣā** (Class.), for which there is no reconstructible Dravidian name.

7.22. Austro-Asiatic languages and Southeast Asia

The following crop names are believed to have originated in Austro-Asiatic languages:

Crop name		*Origin/center of domestication*
(A2)	cotton: OIA **karpāsa** (late Skt)	Southwest Asia
(C6)	foxtail millet: OIA ***kangunī** (late Skt)	China?
(D2)	banana: OIA **kadala** (Epic)	South India–New Guinea
(D3)	betel: OIA **tāmbūla** (Epic)	Malaysia?
(D13)	lemon: OIA **nimbu(ka)** (lex.)	Eastern India
(D16)	pepper: OIA **marīca** (Class.), PSD1 ***miḷ-Vku**	Kerala/SE Asia
(D19)	sugarcane: OIA **śarkarā** (Epic), PD ***cer̲-aku**	New Guinea

For most of these plants, whose domestic origin is to the east of India, this finding is not surprising (see **7.1**). Cotton, however, does not seem to belong to this list, since Austro-Asiatic origin of the name would seem to imply that cotton cultivation originated in Southwest Asia and moved to Southeast Asia before reaching India; yet the oldest evidence for cotton cultivation is from South Asia, as noted in **7.1**(A2). On the other hand, **karpāsa** is not the oldest OIA word for cotton, since it is attested only in the early centuries CE; an older word is **tūla**, attested in that meaning in the Mahabharata, but later replaced by **karpāsa** whose reflexes are found in all the major NIA languages including Sinhala. Since the evidence for the Austro-Asiatic origin of the name **karpāsa** is rather strong,[38] some explanation is required. While the earlier word, OIA **tūla**/PSD ***tū-val**, may have been used for cotton (which presumably came from Arabia/Syria) in the Harappan era, the later word **karpāsa** probably came from cotton producers/users in a more easterly region – which might be anywhere from eastern India to mainland Southeast Asia, where Austro-Asiatic languages are (or were previously) spoken. This linguistic inference needs to be checked against future archaeological findings.

7.23. South Asia and Africa

The following crops are believed to have come to South Asia from different parts of the African continent:

Crop name		*Origin/center of domestication*
(B1)	finger millet: OIA **maḍaka** (lex.), (PD)PSD ***ar/ār-ak**	Africa
(B9)	sesame: OIA **tila** (late V), PSD1 ***eḷḷu**	Ethiopia/South India
(C1)	bulrush millet: OIA ***bājjara**, PSD ***kampu**	Africa
(C10)	sorghum: OIA **yavākāra** (late Skt), **yavanāla** (late Skt) ← PD ***con̲n̲-al**	Africa
(D6)	cowpea: Hindi **lobiyā** ← Pers. **lūbiya**	East Africa
(D14)	okra: OIA **bhiṇḍa** (late Skt) → Ta. **veṇṭai-kā**	West Africa

Regarding sesame, the etymological link between Akkadian **ellu** and South Dravidian ***eḷḷu** suggests that the name (whether originally Akkadian or Dravidian) was probably applied to the crop at the time of Harappan trade with Mesopotamia (see Ratnagar 1981: 80). The OIA **tila** then may well, as in the case of OIA **karpāsa** 'cotton' (see 7.22), be of eastern origin; note Kuiper's proposed Munda etymology (1955).

If sorghum, bulrush millet, and foxtail millet were all taken to South Asia at around the same time (see Simmonds 1976: 92), it is not clear why the OIA name for sorghum was mediated through Dravidian, whereas OIA and Dravidian have distinct names for the other two millets. The early finds of all three crops in South Asia are, not unexpectedly, in the west or the northwest of the subcontinent. One can only assume that different streams of transmission were involved, perhaps overland to northern areas and by sea to peninsular India. Incidentally, the fact that Dravidian created its name for ragi (finger millet) from an old word meaning 'food' suggests that this was a basic crop for that group of Dravidian speakers. In this connection, note Weber's discussion of the importance of ragi at the late Harappan settlement of Rojdi (Weber 1991: 106ff.).

There is linguistic evidence for movement in the other direction as well, that is, from South Asia to Africa. Words similar to the Dravidian word for rice, PD ***var-inc**, appear as **vari** in Malagasy (Madagascar) and as **wari/wali** in various Bantu languages. In addition, the PD ***cīnt(t)-u** 'date' is clearly connected to proto-Bantu **mu-kindu**, though the direction of movement of the word (and the plant) is unknown.

7.24. South Asia and the West

The following European names of crop plants or products have come from South Asia:[39]

(B2) anil (indigo): OIA **nīla** (RV) : PSD1 ***aṇile** → Arabic **an-nīl** → Portuguese **anil**
(B8) rice: PD ***var-inc** → Greek **óryzon**, Arabic **ruz**, Italian **riso**, French **riz**, etc.
(C9) mung: OIA **mudga** → NIA **mūng** → Tamil **mūnku**
(D1) areca: PSD ***aṭ-ay-kkāy** → Malayalam **aṭekka** → Portuguese **areca**
(D3) betel: PSD1 ***ve<u>rr</u>-ilai** → Portuguese **betel**
(D5) copra: PSD ***koppa-rai** → Malayalam **koppara** → Portuguese **copra** 'coconut meat'
(D5) coir: Proto-Tamil ***kai<u>r</u>u** 'coconut fiber, rope made of coconut fibre' → E. **coir**
(D10) bhang: OIA **bhanga** 'hemp' → Hindi **bhāng**
(D10) sunn hemp (Indian/ Bombay/ Bengal hemp): OIA **śaṇa** → Hindi **san**
(D10) ganja: OIA **ganjā** → Hindi **gānjā**
(D13) lemon: OIA **nimbū/limbū(ka)** → Hindi **lemū** → Arabic **laymūn** → OFrench **limon**
(D13) orange: PSD ***nāram-ka** → OIA **nāranga** → Farsi **nārang** → Arabic **nāranj** → Old Provencal **auranj**

(D16) pepper: PSD1 ***tippali** → OIA **pippalī** → Latin **piper**
(D19) sugar: PD ***ceṟ-aku** 'sugarcane' → OIA **śarkarā** → Prakrit **sakkharā** → Persian **śukkar** → Late Latin **succarum** → OFrench **sucre**
(D20) toddy: PD ***tāẓ** → Pali **tāla** → Hindi **tāḍ, tāḍī**

The examples indicate that these words (along with the plants they designate) have been exported to the west at various periods and by different routes. Items (B8) and (D16), for example, were probably carried to the west by Arab traders before the decline of ancient Rome, while some of the other words made the voyage only after the Portuguese contact with India or during the British Raj.

7.3. Summary

This chapter combines botanical and palaeobotanical information about South Asian crop plants with etymological information about plant names in order to illuminate the historical relationships between plants and human societies in South Asia. Section **7.1** presents data on 28 crop plants which have been identified in South Asian archaeological sites (6 from the Early or Pre-Harappan period, 9 from the Mature Harappan, and 13 from the Late or Post-Harappan period), plus 50 additional plants (including 26 trees) for which there is other evidence (linguistic or textual) of early presence in South Asia.

Section **7.2** presents conclusions based on this information regarding prehistoric and historic communication between different language groups and different areas. In **7.21**, it is noted that the majority of cases (eight in all) in which Dravidian crop names were borrowed into Indo-Aryan involve plants whose origin lies to the east of South Asia, suggesting that these plants may have been transported by sea to coastal parts of peninsular India which were dominated (then or later) by Dravidian-speaking groups. Rice may have been among these plants, which suggests that the acquisition of rice in western and southern India may have been independent from its development in eastern India, part of the primary belt of rice domestication. In another group of nine items, the similarities between the Dravidian and Indo-Aryan words seem to point to independent borrowing from the same or related sources, possibly Munda or other Austro-Asiatic languages.

Section **7.22** notes that eight crop names for which Austro-Asiatic etymologies have been proposed designate crops whose origin lies to the east of India, whereas one item (cotton), whose Austro-Asiatic etymology seems secure, probably originated in southwest Asia – a mystery requiring further investigation. Section **7.23** discusses crop plants, particularly millets, which are believed to have originated in Africa, and finds no regular pattern among the etymologies, suggesting the possibility of different routes of transmission to northern and southern parts of South Asia. Section **7.24** lists 15 South Asian crop names which have been borrowed into European languages at various historical periods, beginning with the borrowing of the Dravidian word for rice into ancient Greek and extending into the colonial period.

Notes

1 I use the term 'palaeobotany' to refer to the application of botanical and archaeological knowledge to the study of archaeological plant remains. (See Weber 1991: 13–14 for the meanings and uses of the terms 'palaeobotany', 'archaeobotany', 'archaeopalaeobotany', 'ethnobotany', and 'palaeoethnobotany'.)

2 Prehistorians and archaeologists who deal with South Asian plants have reason to be grateful to Gregory Possehl, the excavator of Rojdi, for including a palaeobotanist on his excavation team.

3 The sources used for Munda reconstructions are Zide and Zide (1973, 1976). Though much additional material is available on Munda, for example in Stampe (2003), considerable Munda expertise is required to make reasonable guesses at the reconstructible forms. For this reason, the materials here are limited to those with published reconstructions. Note the following special characters used in Proto-Munda (PM) reconstructions: X (a laryngeal consonant), R (a reduplicated syllable), V (an unspecified vowel).

4 Krishnamurti's reconstructions are all presumed to be PD (Proto-Dravidian); in some cases I believe they should be placed in PSD (Proto-South Dravidian) or PSD1 (see **8.23**): for example, ***par-utti*** 'cotton' A2(B), ***miḷ-Vku*** 'pepper' D15(B).

5 Zide and Zide note (1973: 7) that there are no reconstructible Munda words for barley, wheat, cotton, ginger, mustard, pepper, coconut, sugarcane, hemp, linen, oilseeds (but see B9), mahua, or maize.

6 Though this is an Indo-European word, it is not clear that its original meaning was 'barley', since the cognates in other Indo-European languages have different meanings (see Southworth 1988: 658). Since this is the only word appearing in the oldest OIA for a grain, it is possible that its meaning was 'grain in general', and that it was applied to barley as an important Indus Valley crop after the arrival of Indo-Aryan speakers in South Asia.

7 There is no word for cotton in the Vedic literature; wool fiber was used for garments and carpets (Randhawa 1980: 296).

8 Sethna (1981) argues that, since the word **karpāsa-** does not appear in Sanskrit until the post-classical period, the whole chronology of Indo-Aryan must be moved back to make that period coincide with the time of the Mature Indus Civilization. This argument has not been accepted by most South Asian prehistorians.

9 I have suggested elsewhere that the very name for the region of Sindh, and the original name of the Indus River (OIA **sindhu**), may be connected with this Dravidian word for 'date'. See Southworth (1988: 659) for further discussion. See also Witzel (1993, 1999b).

10 This post-Rigvedic OIA word appears to be a folk-etymological alteration of some non-IA word, whose original shape may possibly be more closely approximated by OPers **gantum**. See **3.23A**(1) for further discussion.

11 Reflexes are found in the northern and eastern NIA languages, but are missing in the west (Sindhi, Gujarati, Marathi, Konkani).

12 The resemblance between the Dravidian and Munda forms is worth noting, especially as the Munda ***ḍe-** may be a prefix (note the Mundari-Ho form **koḍe**).

13 The fact that the lentil was used in pre-Aryan contexts in South Asia, and that it was proscribed in Vedic ceremonies, point clearly to non-Aryan origin. Dravidian has no attested old words for this crop.

14 The NIA reflexes of this word are found only in central and northwestern NIA. The suggested derivation from **cīnā** 'China' seems doubtful.

15 But note the similarity to the words for 'rice' (item B8).

16 "The possible early age of this culture is suggested by radiocarbon dates obtained from samples of charred rice from the upper levels of Koldihwa giving 5440 and 4530 BC

(uncalibrated). One of the most striking features of these sites is that the Neolithic pottery frequently contains husks of rice. This would seem possibly to be the oldest evidence of rice in any part of the world. However, the early dating is not clearly established: subsequent radiocarbon dates…suggest that the culture may more plausibly be dated to *c.*1600–1500 BC; and the succeeding Chalcolithic phase certainly would support this dating." (Allchin and Allchin 1982: 118)

17 My source for this information is P. Kunhikannan (AKA Kannan Nambiar), an anthropologist who is a native of the area (Cannanore District of northern Kerala). This is an area of very high rainfall, and irrigated land in the region can sustain two or even three rice crops in a year. Kunhikannan notes that in his childhood in the 1940s and 1950s, large areas on the tops and sides of hills were cultivated by the slash-and-burn method with a very simple technology. The vegetation was burned off just before the beginning of the monsoon rains; sowing was broadcast, with a rudimentary hoe used to scratch the surface of the soil, and there was no weeding or irrigation of the land. Only one crop was possible, but often the rice (which was the primary crop) was intercultivated with other crops, particularly millets, pulses, and cucumbers. J. P. Mencher, an American anthropologist who worked in the area in the 1960s, also observed swidden cultivation in this area during her fieldwork. Wild varieties of rice are also found in many parts of Kerala, and possibly in other parts of the west coast.

18 This word seems to have developed from an older Dravidian word meaning 'seed'. See Southworth (1988: 658–9) for further discussion. Witzel (1999b: 26–8) discusses many cognates for this word in other Asian languages, with a possible origin in Southeast Asia.

19 Much has been made of a possible link between items (A) and (B) here; see Southworth (1990b: 227) for further discussion and references.

20 EWA refers to a Telugu **senagalu** and a Parji **cenaya**, both glossed as '*C. arietinum*'. However, these words could not be found in the DEDR.

21 Some NIA reflexes of **atasī** have **l**, for example, Hindi **alsī**, suggesting a northwestern form of Indo-Aryan as the source.

22 According to Weber (1991: 97), all Indian species formerly classified as *Vigna, Dolichos*, or *Phaseolus* are now classified under *Vigna*. (*V. radiata* = *V. radiatus* = green gram = mung; *V. angularis* = *P. mungo* = black gram = urad; *V. unguiculata* = *V. sinensis* = *D. biflorus* = horse gram.) Fuller classifies horse gram as *Macrotyloma uniflorum* (see Fuller 2001a and references).

23 W. M. Steele suggests that *V. unguiculata* (referred to here as 'cowpea') was domesticated in Ethiopia in the fourth millennium BCE, subsequently spreading to West Africa, southern Europe, and Asia, reaching India sometime after 1500 BCE, probably along with sorghum (Simmonds 1976: 183–4).

24 Found in Epic Sanskrit and in Buddhist works, this word has reflexes in all the major NIA languages except Sinhala.

25 OIA **māṣa** 'bean', a Rigvedic word, does not appear to belong here, though Masica glosses its Hindi reflex **māś** as '*P. mungo*' (1979: 80). The OIA and Pali forms are generally glossed simply as 'bean', and the meanings of the NIA forms vary widely.

26 OIA **pūga** has reflexes in Pali and Prakrit only, while **pūgaphala** 'areca nut' (3) has reflexes in Prakrit and western NIA, and ***pūgavṛkṣa** 'areca palm' has a reflex in Sinhala.

27 In Dravidian, though the DEDR has 14 relevant entries, most involve terms for local varieties which are restricted to a single sub-branch of the family. The word cited here, PSD1 ***vāẓai**, is the most widespread term, occurring in 6 SD1 languages. Other words are: ***ar-Vṇṭ̣ṭi*** (Te and Go only) DEDR 205, ***ulu-k** (Pa and Ga only) DEDR 754, ***tāẓ-/tāḷ-/tāḍ-/tāṭ-** DEDR 3181 (Kd, Ki, Kv only).

28 Both these words have reflexes in most of the major NIA languages except Sinhala.

29 OIA **parṇa** 'betel leaf' is originally the general word for 'leaf' (CDIAL 7918).

30 Bloch has suggested a derivation from Dravidian (cf. PSD ***nār** 'fibre from various trees' DEDR 3651), Tamil **kēḷi** 'coconut palm' (not in DEDR; glossed in *Tamil Lexicon* as 'Brahman coconut'; cf. PD ***keṭ-i/a/u** 'tree, bush' DEDR 1941).

31 Pali has **singavera**, and most of the NIA words can be explained as reflexes of this form, though some of the languages of the extreme northwest (known for their conservatism) show indications of an initial ***śr** or ***śṛ** (see CDIAL 12588). Note that Dravidian also uses a word for 'horn' (OIA **śṛnga**) to refer to ginger: cf. Tamil **maruppu** 'horn of a beast, elephant's tusk, part of a lute, branch of a tree, horns of crescent moon, ginger' DEDR 4720).

32 Dravidian has no reconstructible word for okra, but modern Tamil and Malayalam have **veṇṭa**, probably ← IA.

33 This word occurs in the meaning 'berry' in the Atharvaveda (cf. **pippala** 'berry, esp. of *Ficus religiosa*' [= the pipal tree] in the Rigveda).

34 Hindi–Urdu **anār** 'pomegranate' ← modern Persian (Masica 1979: 92).

35 CDIAL notes that J. Przyluski claimed Mon-Khmer origin for this word in a paper in the *Mémoires de la société linguistique de Paris* (XXII: 208). In a later revision of Przyluski's paper (translated into English and published as 'Non-Aryan loans in Indo-Aryan' in Bagchi 1975: 3–34), the author notes that since the publication of his earlier paper "the progress in my research has compelled me to modify it" (Bagchi op. cit. p. 3). In the new version there is no mention of OIA **śarkarā**.

36 An examination of the sources of OIA plant names in **3.24** indicates that only 40 percent of the names can be derived from Indo-European, while the origins of another third are as yet untraced.

37 "Sorghum most probably moved to India from East Africa, initially overland, but movement along the coast must have occurred at an early date, as the dhow trade between the two countries via Arabia is ancient" (H. Doggett in Simmonds 1976: 115).

"It is postulated that *E. coracana* was taken at an early date to India, probably over 3,000 years ago. Sorghum and bulrush millet could have been taken about the same time…It seems likely that it was taken up the seaward edge of the Arabian peninsula along the Sabaean lane" (J. W. Purseglove, op. cit.: 92).

38 Przyluski (in Bagchi 1975: 23–4) cites words for 'cotton' in 11 Austro-Asiatic languages, most of the words consisting of a suffix (of the form **pas/bas/pah** or the like) and different prefixes (or no prefix in some cases), demonstrating that the form of the word is deeply integrated into Austro-Asiatic word structure. While his linking this word to similar words referring to the bow and arrow seems to be a diversion, it is difficult to doubt that the words for 'cotton' are Austro-Asiatic.

39 Another item not mentioned here is **bamboo**, which according to *Webster's Third New International Dictionary* is from Dravidian via Malay (see **3.22A1**).

8

SOME ASPECTS OF DRAVIDIAN PREHISTORY BASED ON VOCABULARY RECONSTRUCTION

Contents

8.1. Introduction

As noted in **1.3**, the field of linguistic palaeontology deals with historical or prehistoric inferences based on ancient (attested or reconstructed) vocabularies. Since Dravidian languages lack the copious ancient literature which is found in OIA,[1] the study of reconstructed forms can provide important information about earlier stages of Dravidian, apart from what is obtained from the study of

linguistic borrowings between Dravidian and Indo-Aryan (**3.22**). The present chapter draws on a collection of Proto-Dravidian (PD) cultural vocabulary derived from Burrow and Emeneau's *Dravidian Etymological Dictionary*, revised edition (DEDR, Burrow and Emeneau 1984). Sections **8.3** and **8.4** present the major conclusions which can be drawn from these materials regarding the habitat, subsistence patterns, and culture of the earliest speakers of Dravidian, and their possible identifications with known archaeological cultures.

8.2. Examination of proto-vocabulary

8.20. Introduction

As described in **2.41**, there are three major subgroups of Dravidian, determined by the sharing of exclusive linguistic innovations: North Dravidian (ND), Central Dravidian (CD), and South Dravidian (SD). According to the assumptions of historical linguistics, any etymon found in two of these groups can be presumed to represent a word in Proto-Dravidian (PD). In the case of Dravidian, two major problems immediately appear:

- The problem of unattested words in the non-literary Dravidian languages (**8.21**), and
- the problem of the relationship between South Dravidian and Central Dravidian (**8.22**).

These two problems lead to a further question:

- the reliability levels of PD and PSD reconstructions (**8.23**).

8.21. The problem of unattested words in the non-literary Dravidian languages

While the SD languages (which consist of two subgroups, SD1 and SD2) are well attested, particularly the four literary languages, those of the other two branches are not. Furthermore, most of the ND and CD languages are in contact with languages of the Indo-Aryan and/or Iranian families, and have absorbed many vocabulary items from these languages, thus eroding their original Dravidian vocabularies. If an etymon is present only in a single branch of a family, however well represented it might be in that branch, it would normally be considered unacceptable to reconstruct it in the proto-language. Thus words found only in South Dravidian, even if present in both branches (SD1 and SD2), do not provide evidence for Proto-Dravidian under the usual assumptions. Krishnamurti, however, has argued that under certain circumstances etyma found only in PSD can be reconstructed for Proto-Dravidian:

> Only four of the Dravidian languages have recorded history and literature starting from pre-CE to the 11th century. The available dictionaries of the literary languages are extensive, running over 100,000 lexical

> items in each case. The [recorded] vocabulary of the non-literary languages is not commensurate… Therefore, most of the cognates turn up in the four literary languages, of which Tamil, Malayalam and Kannada belong to South Dravidian I and Telugu to South Dravidian II. The absence of cognates in the other subgroups cannot be taken to represent the absence of a concept or a term in Proto-Dravidian. The presence of a name (a cognate) in the minor languages and its exclusion in the major languages should lead to a significant observation that the cognate could be lost in the literary languages, but not vice versa.
>
> (Krishnamurti 2003 §1.2.2, reprinted by permission from Bh. Krishnamurti, *The Dravidian Languages: A Comparative, Historical and Typological study*, Cambridge: Cambridge University Press)

Thus, if the likelihood of borrowing from any known language can be ruled out, Krishnamurti would accept words like ***kal-an̲** 'battlefield' (DEDR 1376), ***akaẓ-tt-ay** 'moat' (DEDR 11), ***peṇ-kk-** 'roofing tile' (DEDR 4385) and many others (see Appendix B) as Proto-Dravidian, even though they are only attested in SD1 and/or SD2. While contrary to the accepted canons of comparative linguistics, this argument is not without merit; in order to refute it, one would need to provide evidence that these words came from some other source(s), that is, their absence in the North and Central Dravidian languages is not merely accidental.

As noted in **1.21A**, historical linguists assume that the elements of a language, including its vocabulary, can be accounted for either as inheritances, borrowings, or internal innovations. In this case, the entries in the DEDR show ample evidence to suggest that many of the 'modern' meanings reconstructed by Krishnamurti can be derived from earlier words or meanings: for example, PD ***kōṭṭ-ay** can be reconstructed in PSD with the meaning 'fort, castle, palace' (DEDR 2207a), which is probably derived from a PD word of the same form meaning 'wall', attested in SD1, SD2, and CD (DEDR 2207b). Similarly, PSD ***akaẓ-tt-ay** 'moat' ← PD ***akaẓ-** 'dig, ditch' (DEDR 11); PSD ***pōr** 'fight, war, battle' ← PD ***por-** 'become angry, abuse, quarrel' (DEDR 4540), PSD ***mun-ay** 'battlefield' ← ***mun̲-i-** 'become angry with' (DEDR 5021), PSD ***kaḷ-an** 'battlefield' ← PD ***kaḷ-** 'field, threshing-floor' (DEDR 1376), PSD ***vāḷ** (DEDR 5376) and ***vāy-cci** (DEDR 5399), both meaning 'sword' ← PD ***vā(y)** 'edge, sharp' (see Appendix A, Section G3 and note), PSD ***nāṭu** 'country, district' ← PD ***nāṭ-** 'village' (DEDR 3638), PSD ***ūr** 'town' ← PD ***ūr** 'village' (← 'house'? DEDR 752),***per-kaṭa** 'minister' (Te-Ka) ← PD ***per-** 'to swell, grow big' (DEDR 4411) + ***kaṭa** 'exceeding' (DEDR 1250), PSD ***koḷ-** 'buy' ← PD ***koḷ-** 'bring, take' (DEDR 2151), ***mat̲t̲-** 'barter' ← ***māt̲-** 'change, alter, replace' (DEDR 4834), PSD ***cir-umpu** 'iron' (DEDR 486) ← PD ***ciru-** 'black, charcoal' (DEDR 2552), PSD ***kal-am** 'ship' ← PD ***kalam** 'pot' (DEDR 1305), PSD ***peṇ-kk-** 'tile' ← PD 'potsherd' (DEDR 4385). PSD ***vēnt-ant̲u** shows the

meanings ‘king, Indra, sun, moon, Bṛhaspati’ in Tamil (DEDR 5529), and is probably not to be separated from CD ***vēnt-** ‘god’ (DEDR 5530). The word for ‘palace’ which can be reconstructed as ***kōy-il,** is attested only at the PSD level, though it is composed of elements ***kō** ‘king’ and ***il** ‘house’ which belong to PD; such a compound could have been formed at any time from the PD stage on, but cannot be assumed to have existed in PD. Words for the points of the compass, which can only be reconstructed within South Dravidian, appear to be derived mainly from terms for ethnic groups or topographical features; ‘east’ and west’ refer to the plains and mountains of the Tamil country. (See north in Appendix B, Section C, and the accompanying note.)

PSD ***kī-ṯ-** ‘write’ (mainly in the literary languages) appears to be derived from a PD word meaning ‘line; to draw, scratch’ (see Appendix A, Section I), while ***ōtu** ‘read’ is attested only at the PSD1 level (DEDR 1052) and may have originally meant ‘recite, chant’. PSD ***cat-u-** ‘read’ (this meaning only in Te) has the basic meaning of ‘clever(ness), skill(ful)’ (DEDR 2327).[2] Perhaps surprisingly, ***var-** ‘write’ (DEDR 5263) is attested in this meaning in SD1 (all lgs except Kg), SD2 (Te Go Kd Ki Kw), and CD (Pa Ga). However, given that there is no other reconstructible term to support the notion of general access to written language at the PD level – and considering that speakers of the non-literary languages have come to literacy only in recent historical times – it is more likely that these languages have independently derived the meaning ‘write’ from the older meanings of this word (‘draw, scratch, mark, paint, etc.’),[3] and/or have been influenced by the meanings of cognate words in the literary languages.

At the PD level, there are only two words referring to caste: ***pār** ‘Brahman’[4] and ***māt- ik-** ‘lowcaste person’ (Appendix A, Section F), whereas at least six caste-related terms are reconstructible in PSD or PSD1 (Appendix B, F1). Thus it seems clear that the main development of the caste system took place after the PD period. Some PSD and PSD1 words for caste appear to be derived from words with other meanings, for example names of professions such as weaver (DEDR 2475), basketmaker (DEDR 5092), laborer or scavenger (DEDR 3546);[5] hunter (DEDR 5288), toddy-tapper (DEDR 5049); names of tribal groups such as Kurava (DEDR 1844); negatively evaluative words such as PD ***pol-** ‘base(ness), unclean(ness)’: Ma **pulayan** ‘an untouchable caste’, Br **pōling** ‘stain, stain on one’s character’ (DEDR 4547), PSD ***eḷi** ‘mean, despicable’: Ta **eḷiyaṉ** ‘poor man, man of low caste’ (DEDR 851), PD ***kaṭa-** ‘pass, cross, exceed’: Ta **kaṭaiyar** ‘last, final; men of the lowest caste/status’ (DEDR 1109); or terms referring to high rank or status, like PSD ***mutal-** ‘first’: Ta **mutali** ‘headman’, **mutaliyār** ‘Mudaliyar, member of a landowning caste’ (DEDR 4950), ***vēḷ-** ‘chief’: Ta **vēḷ(ḷ)āḷaṉ** ‘person of Vellala caste’ (DEDR 5545).

These examples make it clear that numerous social and technological developments took place between the PD and PSD periods, suggesting that it would be unwise to unreservedly attribute PSD reconstructed meanings to PD. In other words, the differences between PD and PSD probably represent real differences in the development of the language and the speech community, not mere accidents of attestation.

Such accidents of attestation do of course occur, and therefore any individual item found only in one branch of Dravidian *might* be an isolated remnant of a PD word. Thus PSD words like the following *might* be part of the residue of PD words: PSD ***kappam** 'tribute' (DEDR 1218, attested only in the four literary languages and Tulu); ***ari** 'tax, tribute'[6] (DEDR 216, Ta Ka Te only); ***eṯ-ay** 'chief, lord' (DEDR 527, literary languages only);[7] ***vil-** 'sell', ***vilai** 'cost, price' (DEDR 5421, all SD1 + Te), ***yāḷ-** 'to rule' (DEDR 5157, ditto), ***tēr-** 'car, chariot, temple-cart' (DEDR 3459).[8] Again, the absence of comparable words among the PD reconstructions would suggest caution in attributing these words to PD. In addition, it is necessary to be skeptical about words which occur only in the literary languages (see **8.23**).

Appendix A therefore presents a *minimalist* view of the Proto-Dravidian lexicon, that is, it includes only those items which can be confidently reconstructed according to the standard principles of historical linguistics as described in **1.2** and **1.3**, taking into account the reservations expressed in the preceding paragraphs. Appendix B presents items which can be reconstructed for Proto-South Dravidian, using the same criteria. Section **8.3** discusses the prehistoric implications of these two sets of reconstructions.

8.22. The problem of the relationship between Central Dravidian (CD) and South Dravidian (SD)

The second difficulty alluded to at the end of **8.20** is the ambiguous position of the Central Dravidian languages *vis-à-vis* those of South Dravidian, particularly SD2. Two alternative subgroupings for Dravidian are shown in Figure 2.4 (see Section **2.4**). With regard to these two alternatives, Krishnamurti states: "There is lean evidence to set up a common stage of South and Central Dravidian, but generally a binary division of a speech community is more likely than a ternary" (2003: §11.3). Presumably because of the lack of evidence for a common Proto-South/Central Dravidian stage, Krishnamurti decides in favor of the first alternative, the tripartite division (as depicted on the left side of Figure 2.4).

While it is true that there are no major common innovations shared exclusively by the Central and South Dravidian languages, a look at the details of the differentiation of these two groups provides additional relevant information. Figure 8.1 shows that each major subgroup is defined by a number of common innovations, as indicated by the heavy solid lines in the figure. Common innovations shared by smaller groups within each major subgroup are indicated by lighter solid lines. In addition, there are three lines indicating shared innovations which cross the boundaries of the major subgroups:

(A) development of a perfective participle in ***-c(c)i**, 'shared by the Parji–Ollari–Gadaba subgroup of Central Dravidian with the entire SD2 group;
(B) the change of PDr initial ***c-** to **∅** (zero) in SD1 and Telugu (SD2);
(C) ***okk,** an innovative form of the word for the numeral 'one', shared by Parji–Ollari–Gadaba (CD) with Telugu (SD2). The innovations shown in this diagram are discussed in detail by Krishnamurti (2003, §1.3).[9]

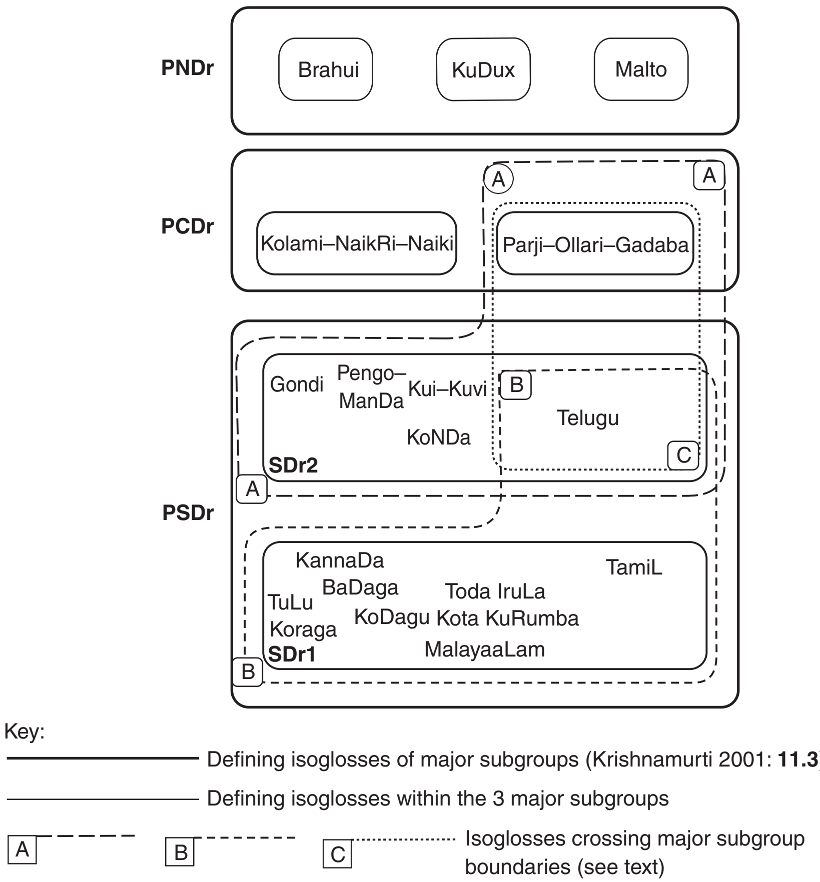


Figure 8.1 Dravidian languages – major isoglosses.

Thus, although there is no basis for assuming a common stage of development for the South and Central Dravidian languages, it is clear that at some stage in the past these two branches were in sufficiently close contact that some innovations could cross the boundaries of the two subgroups. Telugu is still in contact with some of the Central Dravidian languages and with Tamil, and this state of affairs has probably existed for some time. The North Dravidian group, on the other hand, shows no such interaction with the other groups, and thus it is probable that Proto-North Dravidian separated from the PD speech community at a time when Central and South Dravidian were still at least in loose contact. In response to this point, Krishnamurti states:

> Given the geographical location of CD, one would expect several features diffusing in either direction between CD and SD. But we do not find ones that engulf the whole of CD. Note that the number of features that bind

> SD I and SD II is 9 (Fig. 11.3); SD I and SD II also have [respectively] 5 and 7 exclusive innovations. There are four only that are exclusive to CD. The overlapping features are mainly two between SD II and a subgroup of CD: ppl **-cci** and the numeral ***okk**. This is not sufficient evidence to propose a dialect continuum. It does appear that *their geographical contiguity is a recent phenomenon* [author's italics]. Apparently Tuḷu was close to these languages at some point, e.g. the words meaning 'name': Tu. **pudarï**/ Pa. **pidir**; the common plural suffixes ***-ḷ**, ***kVḷ**, etc.
>
> (personal communication, June 2002, quoted with permission)

Thus, while there is no positive evidence for the existence of a Proto-SCD subgroup *in sensu strictu*, there was clearly some diffusion of innovations between the languages of the CD group and those belonging to the SD group. The situation can perhaps be best depicted by a diagram of the type shown in Figure 8.2. Vocabulary items may well have also been borrowed between the two groups, a point which must be kept in mind when reconstructing Proto-Dravidian vocabulary (**8.23**).

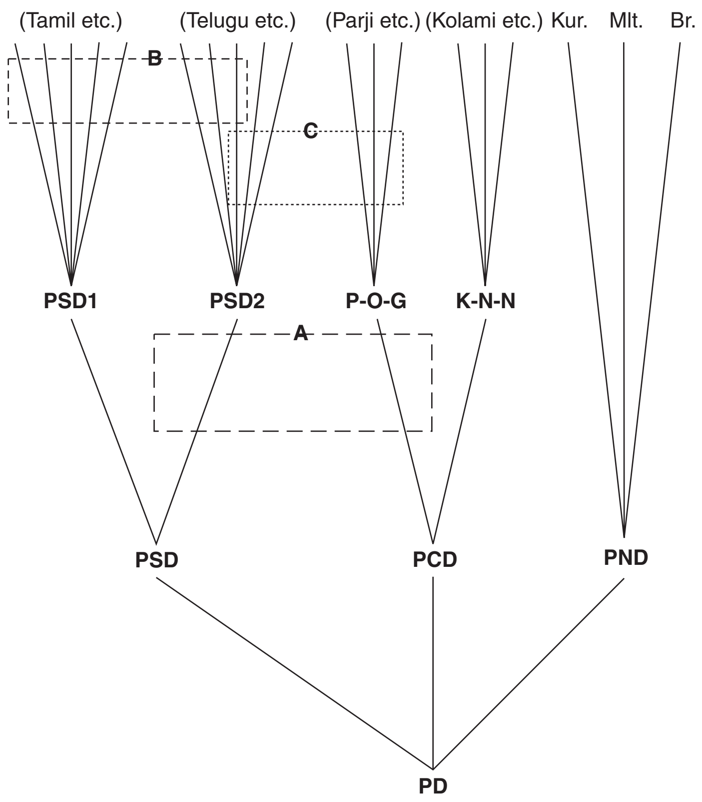


Figure 8.2 Diachronic isogloss map of Dravidian subgroups.

While CD is clearly separate from SD in terms of exclusively shared innovations – the standard criterion for subgrouping (see **1.22B**) – most of the CD languages have been in contact with SD2 languages, especially Telugu, at various times in the past. The fact that some innovations have been (partly) shared between CD and SD2 (**2.41**) is evidence of that; in addition, Krishnamurti (2002) notes that Kolami and Naiki (CD) retain borrowings from Telugu from a very early period. Furthermore, there are cases even among those presented in which CD and SD seem to share a common semantic development which is absent in ND. Some examples are given here:

Reference	*SD–CD meaning*	*ND meaning*	*DEDR No.*
App. A (B)	tamarind (tree/fruit)	sour	2529
App. A (B)	lac tree	kusum tree	4348
App. A (C)	flood	hill-torrent	5503
App. A (D)	cooked rice, porridge, gruel	rice water	0174
App. A (G)	village, town, city	house	0752
App. A (D3)	threshing floor	field	1376
App. A (G4)	thread	to twist (grass etc.)	3726

Apart from this question, those etyma with cognates only in CD and SD2, which have been and are still partially adjacent to each other, must be carefully inspected to rule out the possibility of borrowing or convergent development. Needless to say, etyma with cognates only in *neighboring* CD and SD2 languages are especially suspect. The same would apply to etyma which only have cognates in Kurukh (ND) and a neighboring CD language. Examples:

Reference	*Gloss*	*Attestation*	*DEDR No.*
App. A (D2)	PLOUGHSHARE	SD2: Go; CD: Pa-Ga	4282
App. A (F)	CUSTOM	SD2: Go Ki; ND: Ku	3763
App. A (G3)	BOW	SD2: Go-Pe; CD: Kl; ND: Ku-Mt	1727
App. A (G4)	BASKET	SD2: Kd; ND: Ku	4837
App. A (H)	DEVOTEE/PRIEST	SD2: Ki; CD: Kl	2671

These items are marked with a '?' in the lists in Appendix A, with the comment '{Ctl. India only}'.

8.23. Reliability levels of PD and PSD reconstructions

From what has been said in **8.21–2**, it should be clear that our reconstructions have differing levels of reliability. For PD, the most reliable reconstructions are those with cognates in SD and ND (Kudux–Malto–Brahui), the two branches which have had no direct contact with each other in recent centuries – with the exception of certain types of cases noted before, that is, those subject to possible

borrowing between ND and SD2. Of the 270 items in Appendix A, 100 (37 percent of the total) are of this most secure type.

Perhaps reconstructions with SD1 and CD (Kolami–Naiki–Naikri, Parji–Ollari–Gadaba) are the next best, though the possibility of borrowing between Kannada–Tulu and CD languages cannot be ruled out, as noted previously. Those etyma whose attestations include SD2 (Telugu–Gondi, etc.), which provide the majority of PD reconstructions, are on the whole less reliable than the aforementioned etyma, since they are subject to possible borrowing or influence between SD2 and CD, on the one hand, and between SD2 and SD1 on the other. Telugu is in a pivotal position here, having had close contact with both the CD languages, Kannada, and Tamil.

In general, then, we can say that those reconstructions which are based on SD1 and ND cognates represent vocabulary items which existed during the last period of Proto-Dravidian unity, that is, before the separation of SD, CD, and ND. Reconstructions based on SD and CD cognates, without any ND cognates, may represent vocabulary from this earliest period, or from later times (see **1.31**).

For Proto-South Dravidian (Appendix B), all etyma must be found in both SD1 (Tamil, Malayalam, Kota, Toda, Kannada, Kodagu, or Tulu) and SD2 (Telugu, Gondi, Konda, Pengo, Manda, Kui, or Kuwi). The most secure cases are those which show cognates in the non-literary languages of both sub-branches. When only the literary languages are involved, and especially where the Telugu form closely resembles the Tamil or the Kannada form, the etymology is less secure, since Telugu is known to have borrowed words from both Kannada and Tamil in the past. An example which can serve as a warning is the following DEDR entry:

> **2340** *Ta*. **cappai** a spar of wood placed for the wheels of a car to run smoothly in a sandy road … *Te*. (B.) **cappa** a wedge, a spar of wood placed for the wheels of a car to pass over in a sandy road …
>
> (Burrow and Emeneau 1984:206, reprinted with permission from T. Burrow and M. Emeneau, *A Dravidian Etymological Dictionary*, revised edition, Oxford: Oxford University Press, 1984)

In this case, it is only the presence of the word ‘car’ which provides the clue that this is a modern borrowing from one of these languages to the other (at least the *meaning* is borrowed). In Appendix B, the cases which are judged to be sufficiently secure according to the above-mentioned criteria are glossed in bold type. Of the 171 items in Appendix B, 93 (54 percent) are of the most secure type. Of the remaining 78, 46 (59 percent) are found only in the literary languages, which not only makes them unreliable as PSD reconstructions, but also leaves the possible direction of borrowing unpredictable. Another 30 (38 percent) of these non-secure words show attestation in one or more non-literary SD1 languages, while their only attestation in SD2 is in Telugu. These cases suggest that where borrowing took place between the literary languages, the direction of borrowing was predominantly from Tamil or Kannada into Telugu.[10]

The meanings of reconstructed forms are not precise, in spite of all possible precautions taken in selecting the proto-meanings (see **1.31A**). This is particularly true for the names of faunal and floral species. The meanings of words change over time: for example, the word **corn** means 'maize' in US English, 'grain (cereal plant)' in Britain; similarly, **kaṇak** means 'wheat' in Panjabi, and 'grain of cereal, granule' in other NIA languages. One of the words glossed as RICE in Appendix A (D) probably originally meant 'seed' in PD (see **3.23A(3)** s.v. **vrīhi**). Such changes often take place when a speech community undergoes a change of habitat, or other change which leads to new species moving into the roles of older ones. In a number of cases, the cognate forms refer to different though related species: see, for example, TIGER/PANTHER Appendix A (A4), DEER/ELK/ BISON Appendix A (A8), JACKAL/WOLF Appendix A (A10). Even a proto-meaning like 'donkey' is very imprecise, since it is not clear whether wild or domesticated donkeys (*Equus asinus*) are involved, or even some other species like the hemiones (*Equus hemionus*). Among flora, it is often unclear which sub-species was referred to by the proto-form: see the reconstructions glossed as FICUS in Appendix A(B).

8.24. Early OIA loanwords in Dravidian

This subject has been discussed in **3.22B**, where it was noted that a small number of words can be regarded as possible or probable borrowings from OIA into PD (4 items) and PSD (11 items). In addition, the following may be mentioned as representing possible borrowings from OIA into PSD1 (see Appendix C): ***kacc-ai** 'lower garment' (DEDR app. 20) ← OIA **kakṣyā-, *kīl** 'joint, hinge' (DEDR app. 26) ← OIA **kīla-**. As noted in **3.22B**, most of these potential borrowings consist of everyday words referring to early technical innovations such as axle, axe, needle, platform, crossbar, nail, bell-metal, comb, and cart/wagon.

8.3. Prehistoric inferences based on the reconstructed vocabularies

The vocabularies of Proto-Dravidian (Appendix A) and Proto-South Dravidian (Appendix B) are discussed in the following sections (**8.31–6**) in terms of what they can tell us about the habitat, agriculture, political and economic relations, social relations, material culture and technology, religion, and chronology of these two presumed speech communities. (Appendices A and B are organized into the same main semantic categories, though the sub-categories differ in some cases.)

8.31. Habitat

Proto-Dravidian

Clues to the habitat of the Proto-Dravidian speech community include words for fauna, flora, agriculture, natural phenomena, and (to some extent) technology.

Among domesticable animals PD has words for cattle, sheep/goat, cat, and dog (and, less securely, donkey, horse, and pig); wild animals include reptiles (crocodile, lizard, snake), primates (langur/baboon), various birds (crow, crane, dove, imperial pigeon, peacock tail), and – less certainly – deer, elephant, wild canids (wolf/fox/jackal), felids (tiger/panther/leopard). Unfortunately, we do not know whether the first group of reconstructed words referred to wild or domesticated species, but in general it is likely that they referred to different species or variants at different times and places.

Wild forms of most of these animals (sheep, goat, cattle, chital deer, onager, pig, and probably elephant) have been found in the Indus Valley from the early levels of Mehrgarh in the seventh millennium BCE (Meadow 1986, 1987). Excavations of "late stone age" sites have produced remains of canid, pig, and buffalo (Langhnaj, Gujarat), and dog, *Bos indicus*, buffalo, sheep, goat, and pig, at Adamgarh on the Narbada (Allchin 1979a). Words for aquatic creatures (especially 'prawn' and 'shellfish') suggest a possible awareness of maritime products.

Words for trees include a large number of species native to the South Asian subcontinent, and some – such as the toddy palm (*Borassus flabelliformis*), areca palm (*Areca catechu*), teak (*Tectona grandis*) – which specifically point to peninsular India. The date palm (*Phoenix dactylifera*) was apparently known to the Harappans, and Sindh has long been known as a land of dates (see Southworth 1990b: 228).

Proto-South Dravidian

To these, PSD adds the chicken and the shark (among others), the latter indicating probable knowledge of the ocean and ocean fishing – together with words for 'ocean' and 'seashell' in **3.28C**, and ships and parts of ships in **3.28G5**. Numerous trees are added, including the jack, palmyra, sandalwood, neem, and several citrus species.

8.32. Agriculture

Proto-Dravidian

Terms for crop plants include rice/paddy, onion/garlic, and less securely sesamum, millets (*Sorghum vulgare*, *Panicum italicum*), pulses (*Phaseolus mungo*, *Macrotyloma uniflora*, *Vigna mungo*, *Cajanus cajan*), yam, and sugarcane. The presence of rice, millets, and equids (see above) would allow this society to be included among those participating in a "new agricultural complex" which was widespread on the eastern and southern margins of the Indus Valley by about 1800 BCE, and could have begun considerably earlier, according to Meadow (1989: 68–71). Millets were important in Saurashtra during the Harappan period (Weber 1991), and have been identified at sites in the Deccan and Southern Neolithic sites in the early and middle second millennium BCE (Sankalia 1974: 560, Fuller *et al.* 2001). See further discussion in **8.4**.

The reconstructible terms for agricultural operations make it probable that at least some of the crops mentioned here were cultivated (note the words for sowing and ploughing), processed (note words for husking, winnowing, and grinding), and stored. Some rudimentary methods of water storage and irrigation were available, including tanks and (perhaps) hollowed-out tree trunks used as sluices. Different types of fields were distinguished. In addition to farming, there are words for driving and grazing animals, pointing to the presence of pastoralism. Presumably hunting was still a significant source of food: note NET/SNARE in Appendix A(G2), HUNTING in Appendix A(I).

Proto-South Dravidian

A number of additional crops appear at this level, including a term for ragi (*Eleusine coracana*) – which apparently developed out of a word for 'food/fodder' sometime between the PD and PSD stages – along with perhaps two more millets of uncertain identity.[11] More sophisticated techniques of irrigation are now available, to judge by the new terms for channel, bank, bund, and two new words for sluices. Other important items are words for parts of the plough, seedlings (implying transplantation of paddy and/or other crops?), and a grain measure, which suggests commercial or other transactions involving grain.

8.33. Political and socioeconomic relations

Proto-Dravidian

This speech community is already a society to which the word 'primitive' can hardly be applied. The words for 'king', etc. could of course refer to anything from emperors down to local village chiefs. There seems to be a concept of private property or wealth, along with payment of debts. A word for 'obeisance' reinforces the notion of social hierarchy. Words for 'granary' and 'market' suggest commercial activity. The words for 'Brahman' and 'low-caste person' point to the beginnings of caste. A word for 'barber' suggests the beginning of occupational specialization; note also POTMAKING in Appendix A(G4).

Proto-South Dravidian

At the PSD level we are dealing with an entirely different society, which has new words for kings and chiefs (including one probably borrowed from Indo-Aryan) as well as a word for manager or headman, taxes, armies, divisions of territory, customs or tolls, collection of debts, corvée labor, and perhaps the beginnings of a cash economy. Note that some of these words (e.g. MANAGER, several words for BATTLE, CUSTOMS, TERRITORY, UNPAID LABOR) belong to the less secure category, suggesting that some of these may have been later developments. There are terms for various types of habitations (villages, towns?), castles, forts,

prisons (or storehouses) and palaces, and streets (Appendix B, G1). The caste system appears to be present in well-developed form, along with additional names of occupations, some of which later become caste terms in their turn.

8.34. Material culture and technology

Proto-Dravidian

Words for habitation areas seem to be not yet differentiated from words for dwellings. On the other hand, there are terms for dwelling structures (upper story, stair/ladder, beam) that suggest something beyond simple huts. Words for clothing and adornment are minimal, including only a word for cloth but no specific garments, a few terms for ornaments, and two related to hair grooming. Terms for tools and weapons include basic items like cutting instruments, bows and arrows, digging tools, some metals, and perhaps wheeled vehicles (see AXLE in Appendix A(G4) and the accompanying note). Household items include numerous words for pots, probably originally designating pots of varying sizes, shapes, and functions, as well as baskets and ropes.

Proto-South Dravidian

Terms for urban structures show significant development here, including a number of words for habitations (of different types?) and buildings. Other significant items include many terms for metal objects including weapons and ironwork, wheeled vehicles and ships,[12] umbrellas, garments, and precious stones.

8.35. Religion

For both PD and PSD, the number of words relating to religion is disappointingly small.

Proto-Dravidian

A single word for a deity (SD1 cognates mainly mean ‘king’), two words relating to demonic possession, a word for sacrifice which relates to both hunting and shamanic healing, a possible word for devotee or priest, and a word meaning to bow before or worship.

Proto-South Dravidian

An equal number of new terms: priest/temple servant, another word for a god (‘demon’ in SD1), another term for worship, a word apparently referring to ritual ceremonies involving fire and/or sacrifice, a term for temple (also palace, i.e. ‘god’s or king’s house’), and a word for the cart used to carry idols in temple

processions (probably derived from some sort of early shamanic rituals – see CART/CHARIOT in Appendix B(G3), discussion and note in **8.21**; cf. also Krishnamurti 2003 §1.2.2.2). In both PD and PSD, words for worship and related notions seem to be used with reference to human rulers as well as deities.

8.36. Chronology

The linguistic evidence alone does not provide much basis for determining the chronological limits of the PD and PSD speech communities. The evidence for incipient social stratification, private property, and commercial activity in PD suggests the possibility of contact with state-level societies in the Indus Valley or western Asia, assuming that PD is too early to have had contact with the second urbanization in South Asia during the mid-first millennium BCE. In any case, the terminology involved is not derived from Indo-Aryan. The only possible words of Indo-Aryan origin in PD (see **8.24**) refer to technical innovations of the mid-second millennium BCE or earlier. The evidence of Dravidian loanwords in early OIA (**3.22A**), if acceptable, would also point to the probability of contact between PD and OIA in that period. In fact, if the OIA words (including some Rigvedic words) discussed in Appendix C (see **8.37**) are accepted as loans from PSD1, this would put PSD1 in the mid-second millennium BCE; and since PSD1 is preceded by PSD, which is preceded by PD, the latter might be placed as far back as the third millennium BCE.[13]

8.37. Proto-South Dravidian-1

Though an intensive look at PSD1 would be beyond the scope of this chapter, there are a number of words found only in PSD1 which are historically important. These words, a number of which are listed in Appendix C, have no known cognates in the other branches of Dravidian, including PSD2, and this lack of cognates cannot be blanketly dismissed as due to accidents of attestation. For example, PSD1 ***cakka** 'jack' (tree or fruit) is attested only in Ta Ka Kod, while at the same time there exists a PD ***[*pal-ac/*pan-ac]*** (**8.21B**) as well as a PSD ***coḷ-ay** (**8.22B**) with the same meaning. Such alternate forms are noted in Appendix C where appropriate. Given that the languages of SD1 nowadays cover about twice the geographical area occupied by the rest of the Dravidian languages, and given that this territory includes a large part of the coastline of peninsular India (and possibly included more in earlier times than now),[14] it would not be surprising to find that the speakers of these languages had interactions with members of other groups. The early attestation of some of these PSD1 words as possible borrowings in OIA suggests strongly that the form of Dravidian which was in contact with early Indo-Aryan was PSD1, which of course also had in its vocabulary, in addition to the words mentioned here, all those which have been reconstructed for PD and PSD.

The minimum requirement for inclusion in Appendix C is representation in both of the subgroups of PSD1, namely Proto-Tamil (Tamil–Malayalam–Kota–Toda–Kodagu) and Proto-Kannada (Kannada–Badaga–Tulu–Koraga). In some

cases this requirement has been violated because the items referred to are of likely importance for the prehistory of trade: see the items marked as 'Proto-Tamil' in Appendix C. There are also some cases which have been included in PSD (Appendix B), even though the Telugu cognate is questionable (see **8.23**). Such words, with glosses in Roman (not bold) type and marked with asterisks in Appendix B, might also be considered to belong here. Included in this category are, for example, words for coconut, sandalwood, ocean, north, ragi (*Eleusine coracana*), cardamom, paddyfield, manager, army, scavenger/pariah, ship, cart/chariot, garland, and temple.

8.38. Conclusion of the linguistic discussion

Proto-Dravidian

In summary, the society represented by PD included settled agriculturists who grew and processed rice and millets, kept cattle, sheep/goats, and possibly pigs, and were aware of equids (donkeys, hemiones, or horses), cats, and dogs, as well as wild animals including crocodile, tiger, primates, deer, elephants, and wild canids; they hunted for animals and birds and caught fish. They had some sort of rights to land, along with some social stratification, and possibly the beginnings of caste and occupational specialization. They had a well-developed and varied technology, producing a wide variety of pots, as well as tools for cutting, digging, hunting, warfare (?), and food processing, probably including the use of some metal(s). They built houses or other structures of two or more stories. They used boats and possibly knew some sort of wheel. Their economic system included some form of trade, along with payment of debts and other obligations (contributions to rituals, possibly also fines and/or taxes) and marketing of produce. Their religion included a notion of god and worship, priest or devotee (someone who cares for an idol?), sacrifice (to celebrate a successful hunt and/or for shamanic healing), and (demonic) possession. Thus the community in which this earliest reconstructible stage of Dravidian was spoken was far from a "primitive" society. It must be assumed that there were earlier stages of language (and of technological development, material culture, etc.) which preceded this stage, but they will be beyond our reach unless other linguistic connections come to light.

Figure 8.3 shows the present geographical locations of the various branches of Dravidian, along with the earlier locations inferred from the evidence of place names in Maharashtra (Chapter 9), grammatical features in Marathi, Gujarati, and Sindhi (**1.72B**, Southworth 1974), and Dravidian-type kinship systems in the southern Indo-Aryan languages (**1.72B**, see Trautman 1981). It will be noticed that languages of all three major groups of Dravidian (SD, CD, and ND) are found in eastern central India, in the lower Godavari River system, where Orissa, Madhya Pradesh, Andhra, and Maharashtra adjoin each other. On the assumption that the region of greatest variability is likely to be the oldest region, it is probable that Proto-Dravidian was spoken in the lower Godavari basin, and that it first

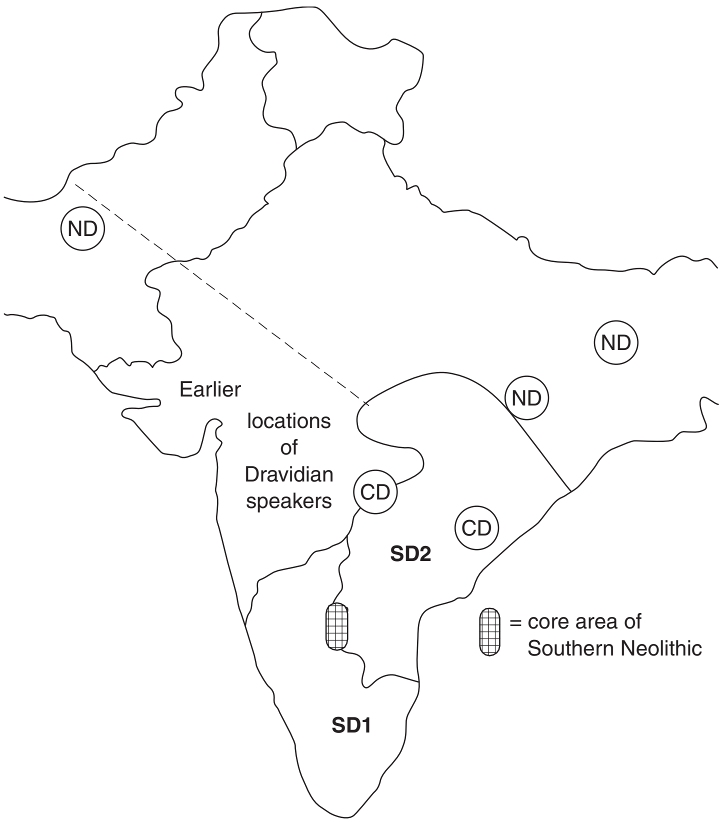


Figure 8.3 Approximate locations of Dravidian subgroups.

began to break up in this same region. Such an assumption would account for the present distribution of Dravidian languages with the least number of required moves.[15] However, Proto-Dravidian may have been spoken in a wider area, for example in Central India or in the western Deccan, which are now occupied mainly by Indo-Aryan languages. Furthermore, other forms of early Dravidian – pre-Proto-Dravidian, or other (at present unknown) branches of Dravidian – may also have existed in these same areas. Thus the suggested location for Proto-Dravidian in the Godavari basin would not preclude the possibility that speakers of an earlier stage of Dravidian entered the subcontinent from western or central Asia, as has often been suggested. Note that this reconstruction is based exclusively on linguistic evidence, namely the current geographical distribution of the linguistic subgroups of Dravidian, along with inferred earlier locations based on loanwords (**3.24**), and may need to be modified when additional (non-linguistic) data are considered, which will be done in the following section.

Proto-South Dravidian

The speech community of Proto-South Dravidian shows new developments in agriculture, including new crops and improved irrigation techniques. A number of professions are now present, including that of weaver, toddy-tapper, and basket-maker, some of them now regarded as castes. This community built forts and palaces, had armies and weapons, land and sea vehicles, and metallurgy. This society was highly hierarchical, socially and politically, with kings, chiefs, and managers, and a developing caste system which included Brahmins, landowning castes, and untouchables, among others. Complex political and economic arrangements are evident, including taxation, customs, debt collection, corvée labor, the beginnings of a cash economy, administrative divisions of territory, and probably a distinction between villages/hamlets and towns/cities. Differences in the attestation of some etyma indicate the probability that the reconstructed words do not belong to a single chronological stage.

On the basis of Dravidian loanwords in (mid-to-late) Rigvedic, which are now accepted by many but not all scholars, OIA–Dravidian contact probably occurred by the mid-second millennium BCE if not earlier. However, without further examination of these loanwords it is difficult to determine which branch of Dravidian was involved in this contact – whether there were unknown branches of Dravidian in the Indus Valley during the second millennium BCE or earlier, or whether a branch of South Dravidian, or even SD1 (e.g. Kannada or a sister language) expanded into that region, perhaps along the west coast of India (see **9.4**).

8.4. Where and when? Putting linguistics and archaeology together

8.41. Proto-Dravidian

The most promising archaeological complex which might be connected with the Dravidian languages is the Southern Neolithic complex, which made its first appearance in the mid-third millennium BCE in the core area shown in Figure 8.3 (the present Gulbarga, Raichur, and Bellary districts of Karnataka, and Kurnool Dt of Andhra Pradesh), and thereafter spread – to judge by similarities in pottery styles, house construction, plant remains, and other features – to “a very vast area from the Krishna–Tungabhadra in the north (or if we include the evidence from Daimabad, on the Godavari) to the Kaveri in the south, and from the Krishna–Godavari mouths in the east to Dharwar in the west” (Sankalia 1977: 142).[16]

The earliest sites of the Southern Neolithic people are found on granite hills, with houses constructed on terraced grounds; “various devices” were used to guide water from one terrace to another. The houses were either round or rectangular one-room structures with low mud walls, the upper part of the wall consisting of bamboo or reed screens, with thatched roofs. They used a “truly amazing”

range of pottery forms (Sankalia 1974: 521).[17] Their tools included stone axes, adzes, choppers, wedges, scrapers, hammers, and lithic blades, and in some sites, copper axes. Querns and grindstones, as well as mortar-like cavities ground into the rocks, were presumably used for processing grains. They created rock-paintings and etchings, including depictions of bulls, deer, gazelles, sheep, goat, horses, peacocks, and serpents; and made ornaments including bead necklaces and ear pendants (?) of shell, semi-precious stones, terracotta, gold and copper (op. cit. 513ff.). A number of these sites are associated with nearby ashmounds, which are presumed to be the remnants of pens for livestock (Allchin 1963). Among animal remains, cattle (*Bos indicus*) predominate, though buffalo (*Bubalus bubalis*), goat (*Capra hircus*), and sheep (*Ovis aries*) are found at numerous sites, while domestic and wild pigs (*Sus domesticus, S. scrofa*), dog (*Canis familiaris*), equines (*E. caballus, E. asinus*, and some unidentified species), deer species, elephant, and rhinoceros remains are present at some sites, along with fish bones (Fuller 2001a: 191; citing Thomas and Joglekar 1994).

Staple food crops consisted mainly of two millets (*Setaria verticillata*, or bristly foxtail, and *Brachiaria ramosa*, or browntop millet) and two pulses (*Vigna mungo*, or black gram, and *Macrotyloma uniflorum*, or horse gram); additional crops which appeared at some sites in some periods included wheat, barley, rice, pigeon pea or toor (*Cajanus cajan*), hyacinth bean (*Lablab purpureus*), wild fruits, oilseeds, some sort of fiber crop, jujubes (*Zizyphus* spp.), figs (*Ficus* spp.), and several additional millets. Finds of tissue fragments lead to the presumption that some sort of tubers were also grown or collected. Flax and cotton were also found, the latter in a late context (Fuller *et al*. 2001: 173–5). Since wheat and barley generally require irrigation, Fuller *et al*. surmise that "the South Indian tradition of tank irrigation and bunding of water near the bases of local hills may have begun, although the start of this tradition is usually attributed to later periods …" (op. cit. 179).

Comparing the preceding with the PD reconstructions presented in Appendix 1, it is clear first of all that there are some positive matches between individual items, such as: cow, sheep, goat, peacock, snake, deer, gazelle, horse, in the animal domain (Appendix A, Section A);[18] rice, sesamum, millets (see further discussion later, this section), horse gram, black gram, pigeon pea, and yam[19] among food crops (Appendix A, D1); the areca palm (*Areca catechu*)[20] and the teak tree (*Tectona grandis*, see Fuller 2001a: 201 with references); words for cultivating and processing grains, words related to pastoralism,[21] and terms related to irrigation (Appendix A, D2–3 and G4); words referring to house construction such as thatch, door, post, screen (Appendix A, G1); earrings and necklaces (Appendix A, G2); generic tools such as axe, adze, chisel, and cutting blade (Appendix A, G3); many terms for different kinds of pots (Appendix A, G4); and a word meaning 'to draw or scratch' (Appendix A, I). Apart from these individual items, the general level of technological and socioeconomic implications of the two assemblages seem to match fairly well: a well-developed agriculture, with some irrigation, accompanied by animal pastoralism and hunting, with a reasonably comfortable lifestyle for the period. Sankalia comments: "These Stone Age folk lived a life

which was in many ways richer and more varied than that of the aboriginals, villagers and many of the town folk today, or during the historical period, stretching back to about 400 B.C." (Sankalia 1977: 526). Fuller *et al*. make the following comment, which can be related to the linguistic evidence for social stratification (see **8.33**):

> The trends towards increasingly diverse cropping systems and some intensification could have facilitated increased populations and/or increased social complexity, and should perhaps be considered as crucial background for understanding later historical developments of more hierarchical societies in South India.
>
> (Fuller *et al*. 2001: 184, quoted with permission from D. Fuller, R. Korisettar, and P. C. Venkatasubbiah,'Southern neolithic cultivation systems: a reconstruction based on archaeobotanical evidence', *South Asian Studies* **17**: 171–87).

On the negative side, there are a number of items which are missing in either the linguistic or the archaeological assemblage. The most glaring gap on the language side is the lack of any words pertaining to burial, a subject which looms large in the archaeological discussions. Another significant lacuna is the absence of words for wheat and barley, though these cereals are not found at all sites (Fuller *et al*. 2001: 174).[22] On the other hand, the linguistic reconstruction contains several words for rice/paddy, which occurs in rather low frequency at only two sites, though its importance increases with time.[23] Also missing from the linguistic inventory are cotton and flax, which do not appear in Dravidian until the PSD stage, though these occur only at the site of Hallur in the later phase of the culture (op. cit. 175). Conversely, many arboreal species are reconstructed for PD which have not (yet) appeared in the archaeological record (Appendix A, Section B).

With regard to the precise species of plants, the linguistic evidence is often ambiguous; this is a particular problem with the millets. The millets reconstructed for PD are ***co<u>nn</u>a-l*** 'great millet, *Sorghum vulgare*' and ***ko<u>t</u>-*** 'Italian millet' (*Panicum italicum, Setaria italica*). Fuller *et al*. have identified two millets – browntop millet and bristly foxtail millet grass – as primary staples of the Southern Neolithic (2001: 174). In order to match these two pairs of millets, we would have to assume that a term which referred to browntop millet (*Brachiaria ramosa*) in the third millennium BCE later came to refer to sorghum, and that a term for bristly foxtail (*Setaria verticillata*) transferred its meaning to *Setaria italica* at some later time. This is not inherently implausible, since *B. ramosa* and *S. verticillata* are cultivated and/or gathered "on only a small or localised scale today" (ibid.), whereas sorghum is a major crop nowadays, occupying over 30 percent of the cropped area in many districts of Maharashtra, Karnataka and Andhra Pradesh (Singh 1974: 162ff., see map p. 165). While *Setaria italica* is not cultivated on the same scale, it is still an important grain crop in India (see **7.17C6**).[24] It would of course be useful to know the words by which these older crops are designated among those who use them nowadays.

Fuller *et al*. have identified six additional millets in the Southern Neolithic, of which three occur only at single sites, and four fail to show clear evidence of cultivation. Three of these occur only in Phase III (1800–1200 BCE).

They are as follows:

Echinochloa cf. colona	(sawa millet)	Phase III only (3 sites), poss. wild
Setaria pumila	—	Phases II and III (2 sites), poss. wild
Panicum sumatrense	(little millet)	Phases II and III (3 sites), poss. wild
Paspalum scrobiculatum	(kodo millet)	Phase III only (1 site), poss. wild
Pennisetum glaucum	(pearl millet)	Phase II only (1 site), imported
Eleusine coracana	(finger millet, ragi)	Phase III only (1 site), imported

Given that these additional millets are not as firmly established at the start of the Southern Neolithic culture as the two discussed previously, it is not unreasonable to inquire whether they might be matched with those reconstructed for PSD – keeping in mind that any of the latter *might* be as old as those reconstructed for PD, or might have come into use at any time between the PD and PSD stages. Following are the reconstructions of words for millets in PSD (along with two from PSD1):

1 PSD ***ārak/*arak-** ‘ragi’ (*Eleusine coracana*)
2 PSD ***ār-Vk-** *Panicum italicum* (Ka Go Kw); *Paspalum scrobiculatum* (Ka Te)
3 PSD ***var-ak-** *Paspalum scrobiculatum* (Ta Ma Ka), *Panicum miliaceum* (Te)
4 PSD ***kam-pu** Italian millet (Ta Ma), *Pennisetum typhoideum* (Ka Te)
5 PSD1 ***tiṉ-ay** *Panicum italicum*[25]
6 PSD1 ***nuv-aṇ-ay** *Panicum italicum*

With the exception of ragi, it is difficult to determine the precise proto-meanings of these words, as the individual languages do not agree in some cases; in addition, there seem to be a large number of synonyms for Italian millet (*Panicum italicum, Setaria italica*). It is also noteworthy that the first three reconstructions are very similar in form, and thus may be ultimately derived from the same source (see FOOD/FODDER in Appendix A, D1). While it may be reasonable to equate reconstruction number 1 with ragi, and either 2 or 3 with *Paspalum scrobiculatum*, the full investigation of these relationships requires more detailed knowledge of the history of agriculture, and agricultural nomenclature, in South India.

Other cases of mismatching between the linguistic reconstructions and the archaeological finds may be mentioned: for example, neither remains nor artistic representations have been found for some of the animals for which terms have been reconstructed (Appendix A, A). Excavation of dwelling structures has not produced

evidence of two-storied structures or stairways/ladders (Appendix A, G1).[26] No wheels or axles (Appendix A, G3 and note) have been reported, though of course if they existed they might have been made of wood and therefore perishable. As to the well-known ashmounds, believed to be cattle-pens (Allchin 1963), the term for 'animal stall' (Appendix A, D2) might be construed to refer to them.

Thus the Southern Neolithic complex is a close, though hardly perfect, fit for the Proto-Dravidian speech community, in terms of specific artifacts, animals, and plant remains. This is not a surprising conclusion, given that Dravidian languages have been dominant in the peninsula since the beginnings of recorded history. The fit between archaeological and linguistic data proves nothing, of course – and yet the chances of such a match occurring are probably quite small, since there is no reason to expect to find a site, or a set of sites, showing a close match in cultural content and location to those reconstructed for a specific proto-language. Thus the fit is perhaps as good as one is entitled to expect.

This does not imply the assumption that the vocabulary of Proto-Dravidian as reconstructed in Appendix A existed at any particular site, or set of sites, at any specific time. Apart from the mismatches between the archaeological and linguistic inventories – some of which might be explained as later diffusions between different branches (e.g. two-storied houses, staircases, wheels), or lexical replacement in one or more branches leading to non-reconstructibility at the proto-stage (e.g. terms connected with burial) – it has already been pointed out (**8.23**) that differences in attestation between different words may imply chronological differences: for example, words which are attested in SD1/SD2 and CD, but not ND, may be later than those found in SD1 and ND. In other words, the assumption of a relationship between PD and the Southern Neolithic implies, at this stage, only that the language which we reconstruct as Proto-Dravidian was probably associated with some of these sites at some time period(s). More research will be needed to make a more precise statement possible. In the meantime, the following may be tentatively suggested:

1 the earliest stage of PD (represented by the items in bold type in Appendix A) belongs to the lower Godavari basin, and can be placed in the early third millennium BCE;
2 the late PD speech community, which included all of the vocabulary of Appendix A, reached from the lower Godavari to the focal area of the Southern Neolithic archaeological complex (Figure 8.3); this involved an expansion of the southern part of the PD dialect continuum, that is, those dialects which later became part of a PSD continuum, a stage reached perhaps by the late third–early second millennium;
3 the beginnings of the Southern Neolithic complex in the mid-to-late third millennium coincide, more or less, with a transitional period in which the three branches of PD became separate speech communities and began to diversify internally, though there was continuing contact between languages of the SD and CD branches;

4 the expansion of the Southern Neolithic between the late-third and late-second millennium is connected with the expansion of the PSD1 speech community into present-day Andhra, Tamilnadu, and Karnataka; PSD2 languages, probably during this same time period, spread northeastward, splitting the original CD community. (Note that the core area of the Southern Neolithic occupies a central position in the present area of the South Dravidian languages, approximately on the border of Kannada (SD1) and Telugu (SD2).)

8.42. Proto-South Dravidian

The preceding suggests that if PSD is connected with any archaeological complex, it would probably be the later phases (perhaps phase 3, 1800–1200 BCE) of the Southern Neolithic archaeological complex. Since this is a wide area, it would imply that what we are calling the Proto-South Dravidian speech community must have shown significant dialectal variation from its earliest stage.[27] However, archaeological explorations of this complex have covered large areas of peninsular India without uncovering traces of the cities or towns, the palaces and fortresses, the metal tools and weapons, or the complex irrigation works[28] predicted by the linguistic reconstructions of PSD. The first societies known to history which might be connected with these reconstructions are the early South Indian kingdoms of the Cholas, Cheras, and Pandyas (Tamil **cōḷa**, **cēra**, **pāṇḍya**).

These kingdoms, which collectively occupied the southern extremity of peninsular India from about the latitude of Kanchipuram southward, are described in the inscriptions of Ashoka (*c*.250 BCE) as "independent neighbors of the Mauryan empire" (Nilakanta Sastri 1964: 8). The first reference to south India in European classical writing is by Megasthenes, who mentions the Pandyas (Pandaea) and Taprobane (an old name for Ceylon/Srilanka), along with various facts and fables about the region (McCrindle 1877: 34). Descriptions of these kingdoms are found in the Sangam literature, a group of poetic anthologies probably composed and compiled in the first three or four centuries CE (Nilakanta Sastri 1955: 55) – though some scholars might place them a few centuries earlier. These descriptions receive substantial corroboration from contemporaneous Mediterranean sources, including writings by Strabo (first century BCE to first century CE), Pliny the elder (first century CE), the geographer Ptolemy (mid-second century CE), along with the *Periplus maris erythraei*, a handbook written by an Alexandrian merchant of the first century CE. Nilakanta Sastri states that "There is perfect concord between the Sangam anthologies, the notices of South India in classical works of the early Christian era, like the *Periplus* and Ptolemy's geography, and the numerous finds of Roman coins of the early Roman Empire in several places in Southern India" (1955: 54).

These sources tell us that the Chola country was divided into a coastal district and an inland district – probably to be identified respectively with the port of

Kāviripūmpaṭṭinam (later Kāvēripaṭṭiṇam) and the inland capital of Uraiyūr (Nilakanta Sastri 1955: 22; Schoff 1974: 241); according to the Sangam texts, these two regions feuded constantly with each other (Nilakanta Sastri 1955: 31). It is not clear at what point sea trade between South India and the west began, but by the time of Ptolemy, trade between Alexandria and India had become quite regular, and the Cholas controlled "the largest and most important shipping" of the Coromandel[29] coast (Warmington 1928: 65). The *Periplus* mentions Indian ships coasting along the shore between the Ganga ports and Damirica[30] (*Periplus* §53). Ptolemy's descriptions make it clear that in addition to regular sea trade between south India and the Roman Empire, the Tamil ports were also the intermediaries for the Roman trade with the far east (Warmington 1928: 128–31). The beginnings of the trade between the Mediterranean and south India can be traced back to the fifth century BCE, when the word **zingíberis** (←PSD ***cinki-vēr** 'ginger-root', see **3.22A(3)** s.v. **śṛṁgavera**) first appeared in Greek (Krishnamurti 2003 §1.2.1.1, note 4),[31] and thus south India may well have been involved in the Indian Ocean coastal trade for some centuries earlier.

The port of Kāviripūmpaṭṭinam, also known as Puhār, is described in Sangam literature as a "celebrated city, full of riches coveted by kings and teeming with sailors," with broad streets and tall mansions, in which different groups such as merchants, brahmans, farmers, physicians, astrologers, and *yavanas* (foreigners) each lived in their own residential quarters. Buildings such as palaces, warehouses, and prisons are mentioned. Descriptions of the overseas trade mention products such as black pepper, pearls, coral and other gems, gold, sandalwood, horses, as well as products from other areas including the Ganga, Srilanka, and Malaya. The government is described as a hereditary monarchy, assisted by ministers and administrators drawn from the wealthy landowning class. Some form of popular assembly existed, but its role is not clear. The royal revenue came from taxes on land, trade (there is a detailed description of the activities of the customs collectors), and conquest. There was a standing army, and wars were frequent and brutal, with little mercy shown to the losers, as well as a standing navy, which engaged in battles on the sea (Nilakanta Sastri 1955: 63ff.).

Agriculture was the chief industry, with much of the fieldwork done by women of the lower classes/castes, who may have been little better than slaves. A clear distinction is made between rich and poor landowners (those who had their land ploughed by others, as opposed to those who ploughed with their own hands). The former often held official posts, and "enjoyed the *jus connubii* with the royal families." Cotton and silk were important products, both in internal and external trade. The ranks of artisans, such as smiths and carpenters, were said to include immigrants from other parts of India, as well as foreign lands. Leather products made by members of the *pulaiya* caste are mentioned. The use of paddy as a medium of exchange is frequently referred to, and it is likely that this was the most common medium in internal trade, metal currency being used primarily for foreign commerce (op. cit. 87ff.).

The other two ancient kingdoms contemporary with the Cholas, those of the Pandyas and Cheras, are also described in the Sangam literature. The Pandyas ruled the southern tip of India, with their capital at Madurai and ports on both the east and west coasts, while the Cheras had their chief port at Kodungallur (anglicized as Cranganore, in early sources also called Muciri, the Muziris of the Greeks), north of present-day Cochin. The Chera capital, Vanji, has not been located with certainty. From a small kingdom on the Malabar coast, the Cheras later expanded along the coast to Kasargod in the north, southward to Cape Comorin, and inland to the Coimbatore region; they had numerous ports on the west coast, several of which have been mentioned in classical sources, and also counted the Laccadive islands in their territory (Nilakanta Sastri 1964: 11–12; Krishna Ayyar 1966: 17ff.). The exploits of the Chera navy – probably consisting mainly of defense against pirates – were celebrated by Sangam poets (Menon 1967: 78).

Nilakanta Sastri, along with other historians, calls the culture of the Sangam age a "composite culture," containing both Indo-Aryan and Tamil elements (1955: 63). This is particularly true in the area of religion, which may partly explain the paucity of words relating to religion, including death rituals, in the reconstructions (**8.35**). Though it is impossible to know how deep the Indo-Aryan component penetrated, the Sangam poets (and presumably their royal patrons and their elite audiences) were clearly aware of the "Vedic and epic mythology of Sanskrit, and the ethical concepts of the Dharmasastras" (op. cit. 1955: 89–90). Words borrowed from OIA are not rare in Sangam Tamil: some examples are **avai** 'assembly' (← ***capay** ← OIA **sabhā**),[32] **atirācan̲** 'superking' (one who has defeated seven other kings) ← OIA **adhirājan** 'king of kings', **kāmakūṭṭam** 'the natural coming together of man and woman' (a hybrid word containing OIA **kāma** 'love, Eros' and PD ***kūṭ-** 'come together' DEDR 1882, Appendix A), **aṅkucam** 'elephant goad, emblem of the imperial Cheras' ← OIA **aṁkuśa**, the word Sangam itself (Ta **caṅkam** ← OIA **saṅgha** 'union'). A famous Chera king was called **Āy Aṇṭiran**, the second part of the name presumably derived from OIA **aṇṭra** 'man, hero' (Nilakanta Sastri 1964: 16). At the same time, the existence of reconstructed PSD words for many of the phenomena described for this period (**8.42**, Appendix B) indicates that much of this development was indigenous, and thus probably prior to the influence of the north Indian Sanskrit-using polities.

Though there is scant archaeological evidence for these early South Indian kingdoms, their existence is supported by the evidence of western classical sources and Roman coins. However, it seems unlikely that the society described in **8.38**, based on the reconstructed vocabulary of Proto-South Dravidian, can be directly related to this period, for the following reasons:

1 *Chronology*: It seems reasonable to place Proto-Dravidian in the early to mid-third millennium BCE on the basis of its probable connection with the Southern Neolithic complex. The following stage, Proto-South Dravidian,

might be expected to follow perhaps 500 or 1,000 years later. On the other hand, there appears to be evidence for the next stage, PSD1, as early as the mid-second millennium BCE (**8.37**, Appendix C). Since the language of the Sangam literature is recognizably early Tamil, and since inscriptions in Tamil occur from the second century BCE onward, the time gap between PSD and the Sangam period is too great.

2 *Level of socioeconomic development*: While the vocabulary of PSD indicates a society which is more advanced (in terms of technology and socioeconomic complexity) than that reflected by the vocabulary of PD and the Southern Neolithic archaeological sites, PSD looks overall like a transitional stage between the agricultural villages of PD and the kingdoms of the Sangam period. The trends visible in PSD, such as the technological innovations in agriculture, the increasing hierarchicalization and social complexity, the growing importance of private property and commerce, as well as the greater role of government and incipient urbanization, all point toward development *in the direction of* the stage represented by the Sangam descriptions, but falling far short of it in many respects. This is especially true if we restrict PSD vocabulary to those words which are highlighted in Appendix B (see **8.23**).

Thus in order to get a clearer picture of PSD society, it is necessary to adjust our interpretation of the PSD reconstructions in two important ways. First, we must not be misled by the reconstructed meanings, which are often heavily influenced by their modern meanings in the literary languages. It was already noted in **8.21** that many of the words for relatively 'modern' phenomena have undergone semantic change from their earliest meanings: for example, PSD **ari** 'tax' has been adapted from an earlier **(v)ari**, which may have referred to some sort of intertribal exchanges of grain and other staple items at an earlier period. Even where all the modern languages agree on a meaning (cf. the words for 'write' in some of the non-literary languages, **8.21**), the reconstructed meaning may represent the end point of a chain of technological or social changes to which the original words adapted independently in the different speech communities.

Second, given that the Southern Neolithic complex expanded fairly rapidly from the original core area, without losing the essential features of the culture, it is likely that there was communication over wide areas within the complex. This would imply that technological, social, and linguistic innovations could pass from one area to another more easily than would be the case with small, isolated groups. As noted in **2.41** and **8.22**, most of the innovations in linguistic structure coincided geographically so as to define subgroups such as PSD1 and PSD2, or smaller subgroups like Proto-Tamil and Proto-Kannada, but even these innovations crossed subgroup boundaries in some cases.[33] Lexical innovations would not be subject to the same constraints, especially if tied to technological or other innovations which had survival value. Thus, as in the case of PD, it is probable that the reconstructed vocabulary of PSD is a composite of various times and

places, rather than the vocabulary of a single uniform group at a particular period of time (see **1.31**). In fact, this may be more true of PSD than of PD, given the larger area it covered. Within this area, over a long period of time, new technological, economic, and socio-political phenomena arose at different times and places, diffusing across a large area from different directions, along with the words used to refer to them.

In spite of the composite nature of the PSD vocabulary, however, it still provides evidence for the continuity of the process of "modernization" in the two millennia which elapsed between the early Southern Neolithic culture in the mid-to-late third millennium BCE, and the rise of the Tamil kingdoms in the mid-to-late first millennium BCE. We know nothing of the antecedents of the Tamil kingdoms. Given the importance of sea trade in their economies, and the early evidence for coastal trade in south India (see earlier), it would not be surprising to find links between such trade and the growth of larger and more complex settlements along the coasts, which have hardly been touched by archaeological exploration as yet. Given the clannish nature of the early Tamil dynasties (Nilakanta Sastri 1964: 13), it is conceivable that some of these early cities grew out of kinship-based alliances between villages.[34]

It must of course be recognized that the descriptions of cities in Sangam Tamil may have involved considerable hyperbole; if we could see them, we might possibly even have some hesitation in using the term 'city' to describe them. As to the ports, there were no facilities in the southern peninsula where large vessels could dock, and they were thus required to anchor offshore and transship their cargoes in small boats, as was done in many of these areas from the time of Vasco da Gama. Nonetheless, there is evidence in Chinese sources for "maritime traders bringing typical Indian products to China as far back as the seventh century BCE" (Nilakanta Shastri 1976: 80), and though these accounts have often been regarded with skepticism, archaeological discoveries in the Philippines, the Malay Peninsula, Java, and Borneo show "clear evidence of a trade contact between the northern Philippines and Southern India running well back into the first millennium B.C." (op.cit. 81). Since the existence of such seafaring ventures implies the presence of shipbuilders, merchants, and financiers, along with professions like carpentry, ropemaking, seamanship, and those required for the production and transport of trade goods – as well as the probable existence of some form of local government – it is possible that the precursors of the Tamil kingdoms go back as far as the early first or late second millennium BCE.

At this stage, it is impossible to interpret the PSD reconstructions, since they point to the existence of a way of life which has not yet been identified archaeologically in peninsular India. This is a case which may be very instructive for historical linguistics, and for the relationship between linguistic and archaeological evidence, but its implications can only be spelled out fully if linguists and archaeologists are willing to collaborate on it. Is it possible that, even with all the precautions taken, the PSD vocabulary projects the state of affairs in the mid-first millennium BCE back a thousand years? If this is the case, then it is important that

linguists recognize that this can happen, and adjust their assumptions accordingly. If, on the other hand, archaeologists do turn up sites which show some of the features implied by the PSD reconstructions, it may be possible to quantify more precisely the length of the time period covered by the reconstructions.

8.5. Conclusions

The Proto-Dravidian speech community was part of a society of the early third millennium BCE which included settled agriculturists who kept cattle and goats, hunted and fished, and knew a large variety of wild animals. They had some sort of rights to land, along with some social stratification, and possibly the beginnings of caste and occupational specialization. They had a well-developed and varied technology, including the use of some metal(s). Their economic system included trade and other economic transactions. Their religion included a notion of god and worship, priest or devotee, sacrifice, and (demonic) possession. In its earliest phase, the PD speech community occupied the lower Godavari basin, and possibly other adjacent areas. Any linguistically related communities which may have preceded this community are not recoverable on the basis of our present linguistic evidence. Possible connections with Elamite, Uralic, or other language groups (see **2.4**, **3.22**) are not ruled out, but are not considered to have been proven.

The Proto-Dravidian speech community was probably associated with the Southern Neolithic archaeological complex of the mid-third millennium BCE from its earliest stages, which may have coincided with the late Proto-Dravidian period. The spread of the Southern Neolithic from its core area was linked with the history of the South Dravidian languages, and probably coincided with the breakup of the PD community into three separate communities (ND, CD, and SD) and the differentiation of these communities into smaller language groups, including the division of SD into SD1 and SD2.[35]

The reconstructed vocabulary of Proto-South Dravidian implies a society which is far more advanced than any found in South India before the beginnings of the late BCE period. There are clear references to some form of government which employs managers, collects taxes and debts, maintains an army, and administers territories of different kinds. There are terms for various types of settlements, buildings, and streets (Appendix B, G1). The caste system appears to be present in well-developed form, along with additional names of occupations, some of which later become caste terms in their turn. New technological items include weapons and ironwork, wheeled vehicles and ships, umbrellas, garments, and precious stones. Even allowing for the probability that some of these terms refer to developments which postdate the proto-stage, these reconstructions seem to anticipate historical and archaeological reality by a millennium or more. The matter is made perhaps more difficult by the likelihood that PSD1, a stage later than PSD, may have existed by the mid-first millennium BCE (**8.37**), suggesting that the date of PSD may have to be placed around 2000 BCE at the latest. This is

a case which may have valuable lessons to offer regarding the relationship between linguistic and archaeological evidence.

8.6. Summary

This chapter examines the vocabulary of Dravidian languages contained in the *Dravidian Etymological Dictionary*, revised edition (Burrow and Emeneau 1984) and reconstructs the approximate forms and probable meanings of words in Proto-Dravidian (PD) and Proto-South Dravidian (PSD). The first section discusses two problems which make it difficult to reconstruct Dravidian: (1) the paucity of lexical materials for the non-literary, as opposed to the literary, Dravidian language, which severely limits the number of items which can be reliably reconstructed (**8.21**); (2) the relationship between the two subgroups CD (Central Dravidian) and SD (South Dravidian), which involved some sharing of structural innovations, although the two branches cannot be considered a single subgroup – implying the likelihood of lexical diffusion between languages of the two groups at all periods (**8.22**). These two factors affect the reliability levels of PD and PSD reconstructions (**8.23**).

Section **8.3** presents prehistoric inferences based on the reconstructed vocabularies, which are given in the appendices to this chapter. The linguistic evidence suggests that the Proto-Dravidian speech community was part of a society of the early third millennium BCE which included settled agriculturists, herders, and hunters; they possessed some form of land ownership, along with some social stratification, and possibly the beginnings of caste and occupational specialization. They had a well-developed and varied technology, including metallurgy. Their economic system included trade, along with payment of debts and other obligations (contributions to rituals, possibly also fines and/or taxes) and marketing of produce. Their religion included a notion of god and worship, priest or devotee, sacrifice, and (demonic) possession. In its earliest phase, the PD speech community occupied the lower Godavari basin, and possibly other adjacent areas. This community was probably associated with the Southern Neolithic archaeological complex from its earliest stages (**8.4**).

The next stage, Proto-South Dravidian, reflects an advanced society with governmental structures including administration, tax collection, and an army, along with various types of habitation areas and urban structures (streets, prisons, palaces). The caste system appears to be present, along with names of occupations. Technology includes many terms for metal objects including weapons and ironwork, wheeled vehicles and ships, umbrellas, garments, and precious stones. Though linguistic evidence (e.g. Dravidian words in OIA) suggests that this stage may be as old as the mid-second millennium BCE, these reconstructions seem to anticipate historical and archaeological reality by a millennium or more. The earliest known communities which might be connected with these reconstructions are the early Tamil kingdoms described in the literature of the Sangam period (early centuries CE). This case may have valuable lessons to offer regarding the relationship between linguistic and archaeological evidence.

Appendix A: vocabulary of Proto-Dravidian

English gloss[1]	*Cognate 1*	*Cognate 2 {comment}*	*Reconstruction*[2]	*DEDR no.*
A. Fauna				
A1. Domestic/domesticable animals				
COW	ā	ND: Ku ōy	*[*ā(m)]*	0334
BULL(OCK)/OX	erutu ‘bull, ox, steer’	ND: Ku addō ‘bullock, ox’	*[*erum-]*	0815
SHEEP/GOAT	(y)āṭu ‘sheep/goat’	ND: Br hēṭ ‘she-goat’	*[*yāṭu]*	5153
SHEEP	koṟi	CD: Nk gorre	*[*koṯ-i]*	2165a
LARGE DOM. ANIMAL	kaṭāy ‘sheep/goat/buffalo’	ND: Br xarās ‘bull(ock)’, xaṛ ‘ram’	*[*kaṭ-ā-]*	1123
CAT (wild/domestic?)	veruku ‘tomcat, wild cat, toddy cat’	ND: Mt berge ‘cat’ {cf. Nahali berko}	*[*ver-uku]*	5490
CAT	pilli	CD: Kl pilli	*[*pill-]*	4180
DOG (calling)	cū ‘sound uttered in setting on dogs’	ND: Br chuh [cu?] ‘call to a dog’	*cū	2718
DOG	nāy (cf. nāḷi, ñāḷi, etc. DEDR 2916)	CD: Ga nēte {cf. Nahali nāy}	*[*naH-]*	3650
DONKEY[3]	kaẓutai	CD: Kl ga:ḍdi	*[*kaẓ-ut-ay]*	1364
?HORSE[4]	SD2: Te guṟṟamu	CD: Kl gurram {see note}	*[*kHuṯṯ-am]*	1711(b)
?HORSE[5]	ivuḷi	ND: Br (h)ulli {see note}	*iv-uḷ-i	0500
PIG (DOMESTIC)	paṉri, vaṉri	CD: Ga panḍ(u)	*[*pan-ṯi]*	4039
A2. Birds				
CROW	kākkai	ND: Br xāxō	*[*kākk-/*kāww-]*	1425
CRANE	kokku	ND: Br xāxūr ‘demoiselle crane’	*[*korV-nk(k)-]*	2125
DOVE	puṟā	ND: Mt pure	*[*puṯ-Vc-]*	4334
IMPERIAL PIGEON	Ma pokiṇa, pōṇa, etc.	ND: Ku poxā	*[*pok-]*	4454
PEACOCK (TAIL)	Ka pīli	ND: Mt pice	*pī-1V/cV	4226
PEACOCK (TAIL)	SD2: Te puri	CD: Kl pūrage	*pur-/pūr-	4367
PEACOCK[6]	maññai, mayil; ñamali, naviram	CD: Ga maŋgil, mayŋil; Kl namli	*[*ñam-V-l]*	2902, 4642
PARROT	kiḷi	CD: Pa kil	*[*kiḷ-]*	1584
?QUAIL	SD2: Ki tumba	CD: Pa tumma {Ctl. India only}	*tump-	3327

(*Appendix A continued*)

Appendix A Continued

English gloss[1]	*Cognate 1*	*Cognate 2 {comment}*	*Reconstruction*[2]	*DEDR no.*
A3. Reptiles				
LIZARD	kauḷi	?ND: Br glūnṭ, klōnṭ	*[*kav-uḷi]*	1338
CROCODILE	mutalai, mucali	ND: Mt boce	*[*mōc-/*moc-V-ḷ-]*	4952
BLOODSUCKER LIZARD	ōtti 'chameleon, bloodsucker lizard'	ND: Mt ute 'bloodsucker lizard'	*[*ōn-tti]*	1053
SNAKE (SKIN)	Ka pere 'skin or slough of a serpent'	ND: Ku cayā̃ per(per)ē 'snake's old skin'	*per-	4417
SNAKE	aravu	CD: Ga rāc	*[*carac-]*	2359
SNAKE	pāmpu	CD: Ga bām(b)	*[*pāmpu]*	4085
SNAKE	SD2: Te pen-jera 'a species of rock-snake'	CD: Pa jēri 'specif. snake' (= Hindi dhāman)	*cer-	2816
PYTHON	mācunam	CD: Pa māyonḍ bām (bām 'snake')	*māc-un-	4793
IGUANA	uṭumpu	CD: Kl uḍug	*[*uṭ-ump-]*	0592
A4. Felidae				
TIGER	pul(i)	CD: Kl pul	*[*pul-i]*	4307
TIGER/PANTHER	uẓuvai 'tiger'	CD: Pa ḍū (Ga ḍū 'panther')	*[*uẓ-uv-]*	0692
CAT(?)	kaṭuvan 'male monkey, tomcat'	ND: Mt gaḍo 'wildcat'	*kaṭ-	1140
(WILD?)CAT	pākkaṉ 'cat; wild cat'	CD: Nk bagale 'cat', Ga bāvki 'wild cat'	*pāk(k)-	4106
A5. Primates				
LANGUR/BABOON	muci 'langur'	ND: Mt muge 'baboon'	*[*muy-cc-]*	4910
MONKEY	kuranku 'monkey, ape'	CD: Kl ko·ti 'redfaced monkey'	*[*kor-…]*	1769
A6. Rodents				
MOUSE	cuṇṭān	ND: Ku coṭṭō	*co(ṇ)ṭṭ-	2661
RAT	eli	CD: Kl elka	*[*el-i]*	0833
SQUIRREL[7]	CD: Kl siḍḍe	ND: Ku ciḍrā {Munda?}	*ciṭ-	2518
A7. Aquatic creatures				
CARP	Ma keṇṭa	ND: Ku kiṇḍō	*[*kiṇṭ-]*	1947
FISH (GEN. TERM)	mīn	ND: Mt mínu	*[*mīn]*	4885
FISH	kayal, cēl 'carp'	CD: Kl kaye 'fish'	*[*kay-V-…]*	1252

FROG	Tu paranṭu ‘bullfrog’	CD: Kl pande ‘frog’	*[*par-Vṇṭu]*	3955
CRAB	ñaṇṭu ‘crab, lobster’	CD: Kl eṇḍe ‘crab’	*[*ñaṇṭ-]*	2901
PRAWN	irā, iravu	CD: Pa reḍā	*[*eṯ-]*	0517
SHELLFISH(?)	Ma kakka ‘shell, cockle’	CD: Kl kāke ‘shellfish, shell of snail’	**kākk-*	1423
A8. Ruminants				
DEER/ELK/BISON	kaṭamāṉ ‘bison’, kaṭam(p)ai ‘elk’	ND: Ku kā̃ṛsā ‘male bādō deer’	*[*kaṭ-]*	1114
DEER/NILGAI	Go koḍal ‘kind of deer’	CD: Pa guḍva ‘nilgai’	**koṭ-*	1664
DEER	uḻai	CD: Pa uṛup ‘spotted deer’	*[*uẓ-u-pp-]*	0694
DEER	Ma mān	ND: máku ‘stag’	*[*mā-y-]*	4780
A9. Insects				
LOUSE	pēṉ	ND: Mt pénu	*[*pēn-]*	4449
NIT	īr(ppī)	ND: Ku cīr	*[*cīr-pp-]*	2625
HORNET	kaṭampai	CD: Pa gaṛdid	*[*kaṭ-]*	1117
SCORPION	tēḷ	ND: Br tēlh	*[*tēḷ]*	3470
A10. Other				
MONGOOSE	mūnkā	CD: Kl muŋgus	*[*mūnk-ūc-]*	4900
PORCUPINE	ey; SD2: Go oy	CD: Kl śēd	*[*cey-t-/*coy-t-]*	2776, 2852
BEAR	enku, elu, uḷiyan	ND: Mt eju	*[*eḷ-V-ñc-]*	0857
ELEPHANT	(y)ānai	CD: Kl enāgī	*[*yA:nay]*	5161
FOX/JACKAL	nari ‘jackal; tiger (cant)’[8]	CD: Kl nakka ‘fox’	*[*nari(-kkV)]*	3606
JACKAL/WOLF	tōṇṭān ‘wolf’	ND: Br tōla ‘jackal’	*[*tōẓ-, *tōẓ-nt-]*	3548
B. Flora: trees				
DATE PALM, *Phoenix dactylifera*	īntu ‘date’[9]	ND: Ku kīndā	*[*cīnt(t)-]*	2617
INDIAN HORSERADISH TREE, *Moringa pterygosperma*	murunkai	CD: Pa muŋga, mulŋga ND: muŋgā ‘a shrub…’	*[*murun-kāy]*	4982

(*Appendix A continued*)

Appendix A Continued

English gloss[1]	*Cognate 1*	*Cognate 2 {comment}*	*Reconstruction*[2]	*DEDR no.*
TODDY PALM, *Borassus flabelliformis*	tāẓ	ND: Mt tálmi	*[*tāẓ]*	3180
TAMARIND, *Tamarindus indica*	(c)intam CD: Kl sintā	ND: Ku citī 'tam. seed', Mt site 'sour'[10]	*[*cin-tta]*	2529
PALAS, *Butea frondosa*	murukku	ND: Ku murkā	*[*mur-ukk-]*	4981
ARECA NUT/PALM	pākku	CD: Kl pōke 'betel'	*[*pānkk-]*	4048
CADAMBA, *Anthocephalus cadamba*	kaṭamp-u/-am	CD: Ga kaṇapa	*[kaṭa-(m)p-]*	1116
COOMB TEAK, *Gmelina arborea*	kumiẓ, kūmpal	CD: Kl kumre	*kūm(p)-iẓ	1742
JUJUBE, *Zizyphus jujuba*	iratti, etc.	CD: Kl rēŋ(g)a {Mt ilkru 'wild plum'}	*[*ira-tt/nt/nk-]*	0475
MARKING-NUT TREE, *Semecarpus anacardium*	cēr	ND: Ku kīrō	*[*kīr-]*	2005
PIPAL *(Ficus infectoria, F. religiosa)*	cuvi 'white fig, *F. infectoria'*, cuvalai 'pipal'	CD: Kl juvi 'pipal, *F. religiosa'*	*[*cuv-]*	2697
PIPAL, *Ficus religiosa*	aracu	CD: Kl ra vi	*[*ar-ac-/-al]*	0202
PIPAL, *Ficus bengalensis?*	āl *'F. Bengalensis'*	CD: Pa ēl 'pipal'	*[*āl-]*	0382
LAC TREE, *Schleichera trijuga*	pū(vu/vam/vatti)	CD: Pa-Ga puy {ND 'kusum tree'}	*[*puc-/*puy-]*	4348
FIG	Ka mēḍi 'glomerous fig tree, *Ficus racemosa'*	CD: Kl mēṛi *'F. Glomerata'*	*[*mēṭi]*	5090
FIG, *Ficus glomerata*	SD2: Ki tōga	CD: Pa tōy	*[*tōnk-]*	3537
DATE, *Phoenix dactylifera'* (fruit)	Ka uttutte, Go uccūṭī	CD: Kl uśti māk	*[*uc(c)-utt-]*	0620

TEAK, *Tectona grandis*	tēkku	CD: Kl ṭek	*[*tēnkk-]*	3452
MAHUA, *Bassia longifolia*	iruppai	CD: Kl ippā	*[*ir-upp-a]*	0485
BAMBOO	vetir, vayir, vēral	CD: Kl vedur	*[*vet-Vr-]*	5485
Terminalia chebula, INKNUT TREE	kaṭu	CD: Pa karka	*[*kaṭu-kkāy]*	1134
EMBLIC MYROBALAN, *Phyllanthus emblica*	nelli	CD: Pa nella	*[*nel-]*	3755
EMBLIC MYROBALAN	SD2: Te usirikā	CD: Kl usurkā 'gooseberry'	*[*uc-Vr-kkāy]*	0574
Terminalia tomentosa	marutu, marutam	CD: Kl madgi	*[*mar-Vt-]*	4718
MANGO, *Mangifera indica*	mā, mānti, mānkāy	CD: Kl māmḍī	*[*mām-]*	4782
MANGO	SD2: Go markā	CD: Pa medi	*[*maṯ-kāy]*	4772
SAL	SD2: Ki jargi	CD: Pa caruŋ	*[*car-]*	2420
JACK, JACKFRUIT	palavu, palā	CD: Pa penac	*[*pal-ac/*pan-ac]*	3988
Eugenia jambolana	Ka nēṟal (etc.)	CD: Pa ñāḍi	*[*ñā(n)ṯ-]*	2917
BELLERIC MYROBALAN	tāṉri, tāṇi	CD: Kl tāṇḍi	*[tānṯ-i]*	3198
C. Habitat				
SHRUBBERY (forest growth)	Ka gumpu 'forest, thick growth'	ND: Ku xoppā 'shrub, bush'	*[*kump-]*	1733
GROVE	cōlai	ND: Mt cále	*[*col-/*cōl-ay]*	2891
HAIL	āli	ND: Mt ali	*[*āli]*	0384
FLOOD/TORRENT	veḷḷam 'flood'	ND: Br bēl 'large hill-torrent'	*veḷḷ-	5503
STREAM (river, brook)	(y)āṟu 'river, brook'	CD: Pa ed 'river, stream'	*[*yA:ṯu]*	5159
MOUNTAIN	kuṉram	CD: Pa kondi	*[*kunṯ-am]*	1864
D. Agriculture				
D1. Crop plants				
FOOD/FODDER	irai 'animal food'	ND: Br iraḡh 'bread, food' {→ PSD 'ragi', App. B, D1}[11]	*[*ār/ar-ak]*	0490

(*Appendix A continued*)

Appendix A Continued

English gloss[1]	*Cognate 1*	*Cognate 2 {comment}*	*Reconstruction*[2]	*DEDR no.*
RICE	Ma maṇṭa 'broken rice, rice dough'	ND: Ku maṇḍī 'cooked rice, meal'	**maṇṭ-*	4679
RICE (← 'seed'?)[12]	(v)ari(ci)	CD: Pa velci	*[*var-inc]*	0215
RICE (grain)	alaku 'grain/ear (of paddy/other grain)'	ND: Ku alkhṛā 'parched rice'	*alk	0255
PADDY	Ka key 'crop; paddy plant'	ND: Mt qesu	*key-(c-)	1936
EGGPLANT/BRINJAL	vaẓutalai/vaẓutuṇai	ND: Mt baṭango	*[*vaẓ-Vt-]*	5301
POD	ēn̲al 'ear of corn'	ND: Mt canu 'pod'	*[*cēn̲-]*	2798
PORRIDGE (rice/millet)	Ka ambali 'rice/millet porridge'	ND: Ku am(b)ṛī 'rice water'	*[*amp-ali]*	0174
ONION/GARLIC	uḷḷi	ND: Ku uli	*[*uḷḷi]*	0705
YAM	cēmp-u, -ai	ND: Ku kisgō 'yam'	*[*kic-ampu]*	2004
KANJI (rice water)	kañci 'rice-water'	CD: Ga genji 'water from boiled rice'	*kañci	1107
SESAMUM	nū(vu)	CD: Ga nuvul	*[*nū(v)-]*	3720
SORGHUM VULGARE	con̲n̲al	CD: Kl sonna **{see 7.13(C)10}**	*[*con̲n̲a-l]*	2896
ITALIAN MILLET	kural	CD: Pa koyla	*[*kot̲-]*	2163
GRAM	payar̲u *'Phaseolus mungo'*	CD: Kl pesāl 'green gram'	*[pac-Vt̲/Vl]*	3941
HORSE GRAM	koḷ	CD: Pa kol	*[*kol]*	2153
BLACK GRAM	uẓuntu	CD: Kl urunde	*[*uẓ-untu]*	0690
BLACK GRAM, *Phaseolus mungo, P. radiatus*	SD2: Te minumu	CD: Ga mindil	*[*min-]*	4862
HEMP	SD2: Go boyal 'sunn hemp'	CD: boili 'hemp'	*[*boy-Vl]*	4535
PIGEON PEA, *Cajanus indicus*	tuvarai	CD: Kl togar	*[*tu-var-]*	3353
PIGEON PEA, TUR PULSE	SD2: Te kandi	CD: Ga kardi	*[*kar-Vnti]*	1213
SUGAR CANE	Ka ceruku	CD: Kl saragurak	*[*cet̲-]*	2795
WHEAT/RICE	SD2: Kd kuli 'paddy'	?ND: Br xolum 'wheat' {cf. **godhūma, 3.23(1)**}	*[*kūl-i]*	1906

D2. Agricultural operations and implements

DRIVE (animals)	To mīṟ- 'drive (buffalo) on migration'	ND: Br miṛing 'drive away, drive (area) for game'	*[*mīṯ-]*	5593
GRAZE	mēy-'graze, feed'[13]	ND: Br bei 'grass for grazing'	*mēy-	5093
TO PLOUGH	uẓu-	ND: Ku uinā (ussas)	*[*uẓ-V-]*	0688
PLOUGH	araka 'plough w/bullocks'	ND: Mt are 'plough'	*[*ar-V-]*	0198
PLOUGH	ñāñcil, nāñcil	CD: Kl naŋgli {see **3.21, 3.22A1**}	*[*ñāṅ-kVl]*	2907
PLOUGH (TEAM)	ēr 'plough, yoke of oxen'	CD: Pa cer 'team of bullocks' {cf. 2814 'join'}	*[*cēr]*	2815
?PLOUGHSHARE	SD2: Go puyil	CD: Pa puyil {Ctl. India only}	*[*pu-y/nk-]*	4282
FURROW	cāl	CD: Ga sāla	*[*cāl]*	2471
SEED/SOW	vittu 'seed, sow'	ND: Mt bīci 'seed'[14]	*[*vit-]*	5401
REAP	koy-	ND: Mt qoye	*[*koy]*	2119
SICKLE	SD2: Go heṭṭaṛ	CD: Pa cetal	*ceṭṭ-aḷ	2756
GRIND (grain)	nūṟu-'crush'	ND: Br nusing 'crush, grind'[15]	*[*nūṟ-]*	3728
WINNOW	teḷi-	ND: Ku telnā	*[*teḷ-]*	3435
WINNOW	nēmbu	ND: Ku nābnā	*[*nē(m)p-]*	3769
WINNOWING BASKET	SD2: Te cēṭṭai	ND: Mt kétnu	*[*kēṯṯ-]*	2019
TO CHURN	kaṭai	ND: Mt gaṭye	*[*kaṭ-]*	1141
WEED/PLUCK	kaḷ	ND: Br xalling 'uproot, gather (vegetables, grass for fodder)'	*kaḷ-	1373
HUSK OF GRAIN	umi	ND: Ku uŋk	*unk	0637
TO HUSK (GRAIN)	CD: Kl īk-'pound (grain)'	ND: Ku īkhnā 'half-husk'	*[*īk-]*	0535
CHAFF	poṭṭu	CD: Ga poṇḍ	*poṭṭ-/*poṇṭ-	4491
GRAIN RECEPTACLE (STORAGE)	Ka komme	CD: Nk ghummi	*komm-	2117
STALL (FOR ANIMALS)	toẓu 'cattle-stall, manger, pound…'	CD: Pa cakur-tol 'cattle-stall' (cakur 'bull', DEDR 2267)	*[*toẓ-]*	3526
HERD/FLOCK	mantai	CD: Pa manda	*mant-ay	4700
SLUICE/DRAIN	tūmbu 'tube, sluice, channel…'	CD: Pa tum bukka 'tree trunk used for draining water from fields'	*[*tūmpu]*	3389
TANK	ceṟu	CD: Kl ceru	*[*keṯ-ayl-uvu]*	1980
?WELL	SD2: Te nuy(y)i, nuvi	CD: Kl nuvvi (← Te?)	*[*nūy]*	3706

(*Appendix A continued*)

Appendix A Continued

English gloss[1]	*Cognate 1*	*Cognate 2 {comment}*	*Reconstruction*[2]	*DEDR no.*
FISHHOOK	kāḷam (for alligators)	CD: Nk gaḷ	*kāḷ-	1495
COWBELL	SD2: Go hirna	CD: Pa jinna	*ciṛ-nk-	2515
DIGGING TOOL	SD2: Go kūnj 'hoe, digging stick'	CD: Ga gusi 'crowbar'	*kū(ñ)c	1879
?CLEAR LAND[16]	SD2: Go erk 'cut down (grass, bushes)'	CD: Pa erk 'cut down bushes to clear land'	*eru-kk-	0824
D3. Types of land				
LOW-LYING LAND	Ka kudaru, kudiru, kuduru 'low ground, streambed'	ND: Ku kudar 'low ricefields'	*kut-Vr-	1700
UNCULTIVATED LAND	karu-nilam 'barren soil, wasteland'	ND: Mt qawru 'forest, jungle'[17]	*kar-V	1285
UNCULTIVATED LAND	SD2: Go kaman 'forest'	ND: Mt kambane 'uncultivated land'	*kam(p)-an-	1228
FIELD	pulam 'arable land, rice field'	CD: Kl polam 'field'	*[*pol-am]*	4303
FIELD	cey 'field, esp. wet field'	CD: Kl kēn 'field' {ND 'mud'}	*[*key-m]*	1958
PADDYFIELD	vayal	CD: Pa vāya[18]	*[*vāy-/vay-]*	5258
THRESHING-FLOOR	kaḷam, kaḷan	CD: Kl kalave {→ PSD 'battlefield', App. B, E}	*[*kaḷ-am/-an]*	1376
E. Political and economic relations				
KING/LORD	kō(n) 'emperor, king'[19]	ND: Ku kōhā'great, high-ranking' {cf. 2178 Ta kō 'mountain'}	*[*kō(n)-]*	2177
CHIEF[20]	vēḷ 'chief, petty god'	ND: Ku bēlas 'king, zamindar, god'	*vēḷ-	5545
KILL	ceru 'kill, destroy'[21]	ND: Mt qerme 'destroy, massacre'	**ker-*	1981
PAY (TAX/FINE?)	iru 'pay (tax/debt)'	ND: Ku erkhnā 'lose money, be fined'	*et-(kk)-	0521
SELL	virru/vira 'sell'[22]	ND: Mt bise	*vit (t)-	5459
PAY DEBT	moy 'payment of obligation or contribution'	ND: Mt múje 'pay debt'	*moc-/*moy	5121
DEBT	vāram 'tenant's share of crop to be paid'	CD: Ga vāri 'debt'	*vār-	5359
STEAL/THEFT	kaḷ 'rob, steal', kaḷḷan 'thief'	ND: Br xalling 'lift (cattle)'	*[*kal]*	1372
THIEF	tonkan 'thief'	CD: Kl ḍonga	*toṅk-an	2982

PROPERTY/WEALTH	uri-mai ‘ownership, possession’[23]	ND: Ku urb ‘rich’; urbas ‘master, landlord’	*ur-i/p(p)-	0654
PROPERTY/WEALTH	uṭai ‘wealth’, uṭaimai ‘wealth, ownership’	ND: Ku uṇḍrī ‘mistress of house’	*u(ṇ)ṭ-	0593
GRANARY	Tu. kadike	CD: Pa gadeya	*kat-ik(k)-	1192
MARKET	ankāṭi[24] ‘bazaar, bazaar street’	CD: Kl angaḍi ‘bazaar’	*aṅ-kā-ṭi	0035
LABORER’S WAGES	kūli	CD: Pa kūli	*kūli	1905
F. Social relations[25]				
HOUSE(HOLD)	Ma paḷḷi ‘hut’	ND: Ku erpā-palli ‘household, family establishment’ (erpā ‘house’) {cf. App. B, G1}	*[*paḷḷi]*	4018
?SERVANT[26]	āḷ	ND: Ku āl	**āḷ*	0399
OBEISANCE	Ka moẓgu ‘bow (to another)’	ND: Ku mulkhrnā ‘kneel’	*moẓ-k(k)-	5123
TO BEG, BEGGAR	tevvu ‘beg’, Ma tēra ‘beggar’	ND: Ku tembnā ‘beg for alms’, tembᵃrus ‘beggar’	*te(m)p-	3431a
BRAHMAN	pār	CD: Nk pār	*pār-	4091b
BARBER/SHAVE	ciraiyan/cirai	ND: Mt qeru/qére	*ker-	1564
LOWCASTE PERSON	Ka. mādiga ‘cobbler, outcaste’	CD: Nk madig ‘man of sweeper caste’	*māt-ik-	4810
MARRIAGE	maṇam ‘union, marriage…’	ND: Mt manye ‘to love, marry’	*maṇ-	4667
MARRIAGE	CD: Ga ōdur	ND: Ku uinā ‘to marry’	*ōt-	1051
WIDOW	kaintalai, ka(y)iṉi	CD: Pa Ga kēṭal	*kay-nṯ-al	2028
?CUSTOM	SD2: Go nēng	ND: Ku nēg ‘ceremony, sacrificial rite, received custom’ {Cent. India only}	*nē(n)k	3763
G. Material culture and technology				
G1. Habitations				
TO THATCH	vēy-	ND: Br bērifing ‘clothe, thatch’	*[*vēy]*	5532
WALL	Ka gōḍe	CD: Kl goḍā {→ PSD FORT/CASTLE, App. B, E}	*kōṭṭ-	2207b
HABITATION[27]	ūr ‘village, town, city’	ND: Br urā ‘house, wife’ {→ PSD TOWN, App. B, G1}	*[*ūr]*	0752

(*Appendix A continued*)

English gloss[1]	*Cognate 1*	*Cognate 2 {comment}*	*Reconstruction*[2]	*DEDR no.*
HABITATION[28]	kuppam 'village, hamlet'	ND: Ku kumbaa 'log hut'	*kumpp-	1732
HOUSE	il 'house, home, place, wife'	CD: Kl ella 'house'	*[*il]*	0494
UPPER STORY	aṭṭam 'terraced roof, upper story'	ND: Ku attā 'raised platform, loft'	*aṭṭ-	0093
UPPER STORY	Ma māṭam 'house with upper story'	CD: Kl maˑḍi 'upper story of a building'	*māṭ-	4796a
BEAM	paṭṭai 'palmyra timber, rafter'	ND: Ku paṭṭā 'beam in oil mill'	*paṭṭ-ay	3875
STAIR/LADDER	Ka kūlu	CD: Pa kūl	*kūl-	1908
DOOR	CD: Ga val 'door'	ND: Ku balī	*val-	5277
POST	Ka muṇḍige	CD: Kl muṇḍe	*muṇṭ-	4948
SCREEN/MAT/THATCH	taṭṭi 'screen, mat (as of grass/rattan)'	CD: Ga taṇḍaka 'bamboo curtain'	*taṭ-	3036
G2. Clothing and Adornment				
FOOTGEAR WITH STRAPS	ceruppu 'sandal'	ND: Ku kharpā 'ankle straps'	*[*keruppu]*	1963
CLOTH	Go gisīṛ	ND: Ku kicṛī 'cloth, garment'	*kic-Vr-	1521
COMB/TO COMB	vāri/vār-	ND: Ku bāgirkā/bāgnā	*vār-kk-	5357
CHIGNON	koppu	ND: Ku xoppā	*kopp-	2110
BRACELET	SD2: Go sinnūm	CD: Ga cinup	*cin(n)-u-	2531
BELL	SD2: Te mū̃ga 'small bell or tinkling ornament'	CD: Kl muŋge 'anklet bell'	*mūnk-	5025
EARRING	kuḻai	CD: Kl kuḍka	*kuḻ-kk/nk-	1823
NECKLACE	kāṟai 'gold/silver collar'	CD: Nk kara 'silver necklace'	*kāṯ-ay	1506
ORNAMENT	muruku 'ear ornament'	CD: Ga murug 'bracelet'	*mur-uk-	4979
INK/COLLYRIUM (←SOOT?)	mai 'collyrium, etc.'	ND: Ku maĩs 'ink' {Nk mas, Go masi 'soot'}	*may-(c)-	5101
G3. Tools/Weapons				
AXE	maḻu	ND: Mt masu	*[*maḻ-V]*	4749
KNIFE[29]	vāy 'mouth; . . . edge as of knife'	ND: Br bā 'mouth, aperture, edge of knife'	*vā(y)	5352
ADZE	cīy- 'cut with adze'	ND: kislā 'adze'	*kī(y)-	1612
BOW	vil	ND: Br bil	*[*vil]*	5422

?BOW	CD: Kl gunti	ND: Ku guṛthā, gunthā {Ctl. India only}	*ku(n)t-	1727
SHOOT ARROW(S)	ey-	ND: Ku ínjnā	*e(n)c-	0805
ARROW	kōl 'stick, … arrow'	CD: Pa kōl 'shaft of arrow'	*[*kōl]*	2237
ARROW	ampu	CD: Pa amb	*[*ampu]*	0178
DIGGING TOOL	kuntāli, kuntāḷi 'pickaxe'	ND: Mt qodali 'spade'	*[*kun-tāl]*	1722
DIGGING TOOL	pārai 'crowbar, hoe'	CD: Ga pāra 'spade'	*[*pār-ay]*	4093
DRUM	tampaṭṭam	CD: Nk tappa 'drum, tambourine'	*tapp-/*tamp-	3082
IRON/WEAPON	irumpu 'iron, weapon'	CD: Kl inum	*[*cirum-pu]*	0486, 2552
NET/SNARE	valai	CD: Ga val(l)a	*val-	5288
SHOULDER-YOKE	kāvaṭi	CD: Kl kāvaṛi	*[*kā-vaṭi]*	1417
AXLE[30]	irucu	CD: Kl i·rcu	*irucu	0484
BOAT	ōṭam 'boat'	CD: Pa ōḍa	*ōṭ-am	1039
METAL	poṇ 'gold, metal, iron …'	CD: Ku pannā 'iron'	*[*poṇ]*	4570
CHISEL	uḷi	CD: Ga ulli	*uḷi	0699
G4. Household items and activities				
POT	Ka aḍigera	ND: Ku arī	*aṭi-	0075
POT	kōy 'vessel for toddy'	ND: Br xō 'cooking pot'	*kōy	2225
POT	patalai 'lg-mouthed pot'	ND: Mt patli 'cooking pot'	*pat-Vl-	3909
POT	maṭā, miṭā 'lg. earthen vessel'	ND: Mt meṭa 'chatty, pitcher'	*mVṭ-	4651
POT	kalam 'vessel, plate …'	ND: Br kalaṇḍ 'broken earthen pot, any old pot' {→ PSD SHIP, App. B, G3}	*[*kal-am]*	1305
POT	Kr cōra 'milk pot'	CD: Pa cōra 'earthen pot'	*cōra	2886
POT	tūtai 'small earthen vessel'	CD: Kl duttā 'earthen pot'	*tūt(t)-ay	3385
POT	muntai 'small vessel'	CD: Pa mutta 'basin'	*munt-/*mutt-ay	4965
POT	akal, akaḷam	CD: agle	*ak-Vl/Vḷ-	0009
POTMAKING	Ka bān 'make as a potter does'	CD: Pa vāñ- 'make, as a pot'	*vāñ-	5327
METAL POT/CUP	kiṇṇam, kiṇṇi	CD: Ga gina	*kiṇ-	1543
SPOON/LADLE	SD2: Te t(r)eḍḍu	ND: Mt ṭaḍu	*teṭ(ṭ)-	3411
BASKET	taṭṭu 'broad shallow basket'	CD: Ga taṭṭa 'wide basket' {basic mg. 'flat, flat vessel'}	*taṭṭ-	3035

(*Appendix A continued*)

English Gloss[1]	*Cognate1*	*Cognate2 {comment}*	*Reconstruction*[2]	*DEDR No*
?BASKET	SD2: Kd mānika	ND: Mt manxā {Ctl. India only}	*mān-Vk-	4837
BELL	Ko jelk ‘anklet with bells’	CD: Kl zilal ‘small cowbell’	*cil-	2572
ROPE (for suspending articles inside house)	uṟi	CD: Nk uttu	*uṯ(ṯ)-	0708
ROPE, PLANT FIBER	nār	CD: Kl nārā	*nār	3651
PESTLE/TO POUND	Tu ujjeru ‘pestle’	CD: Ga uyup ‘pound (rice, etc.)’	*uc-(up-)	0583
MORTAR/TO POUND	Ka aval ‘pounding/beating in mortar’	CD: Pa cavil ‘mortar’	*cav-Vl-	2391
CURDS	peruku ‘curdled milk’	CD: Kl pereg	*[*peṯ-, pēṯ-]*	4421
TO CHURN	kaṭai	ND: gaṭye	*kaṭ-ay	1141
SALT	uppu	CD: Kl sup {cf. Nahali copo ‘salt’}	*[*cup]*	2674a
FLOUR	piṇṭi	CD: Ga pinḍi	*piṇṭi	4162
FERMENTED LIQUOR	kaḷ ‘toddy, honey’	CD: Kl kal ‘liquor’	*[*kaḷ-]*	1374
TODDY	māli	CD: Pa māl	*māl-	4822
SPIN/TWIST	SD2: oḍuku ‘spin’	ND: Ku ōjnā ‘spin, twist’	*oẓ-uk-	1012
WEAVE	ney-	ND: Ku essnā	*[*nec-/*ney-]*	3745
CHARCOAL	iruntai (cf. iravu ‘night, black’)	CD: Pa ciruŋ (cf. Kl cirum ‘dark’)	*cir-un-	2552
H. Religion				
POSSESSION	pēy ‘devil … possession’	ND: Mt peypeyre ‘feel fervent or animated’ {→ ‘spirit’ in PSD, App. B, H}	*[*pēy]*	4438
DEMON/POSSESSION	Ma cōku ‘demon’	CD: Pa cokk- ‘to possess’ (of spirits)	*cōk(k)-	2870
SACRIFICE	SD2: Go lākāna, lākānj ‘sac. after successful hunt’	ND: Br alēj ‘sac. in which patient witnesses slaughter…’	*(a)ḷV(n)k…	0297
BOW	Ko aḍ mug ‘bow before god/godlike person’[31]	ND: Ku mulkhrnā ‘kneel with head to ground’	*muṛk-/moṛk-	5123
GOD	vēntan̠ ‘king, Indra, sun, moon, Bṛhaspati’	CD: Ga vēndiṭ ‘god’	*vēnt-	5529–5530
?DEVOTEE/PRIEST	SD2: Ki supaṛi ‘holy, devoted’	CD: Kl sup(p)ari ‘diviner, priest’ {Ctl. India only}	*cuppari	2671

I. Miscellaneous

CARRY ON HEAD[32]	cuma 'become heavy, etc.', cummaṭu 'pad for headload'	ND: Ku kum'nā 'carry on head'	*[*cum-V-]*	2677
MEDICINE/POISON	maruntu	ND: Ku mandar 'poison'	*[*mar-untu]*	4719
ASSEMBLE, MEET	kūṭu	ND: Ku xōṇḍrnā	*kūṭ-	1882
SING/SONG	SD2: ēla 'kind of song'	ND: Mt ēle 'sing'	*ēl-	0913
SING	pāṭu	ND: Ku pāṛnā	*[*pāṭ-]*	4065
DAY/SUN/TIME	nēram	ND: Br dē 'sun(shine), day, time'	*[*nēram]*	3774
TOMORROW	nāḷai	ND: Ku nēlā	*[*nāḷ-]*	3656
MOON/MONTH	nilavu, nilā 'moon(light)'	CD: Pa neliñ 'moon, month'	*[*nel-]*	3754
YEAR	(y)āṇṭu	CD: Nk iyāṇḍ 'this year'	*[*yA:ṇṭu]*	5153
LAST YEAR	SD2: Te niruḍu	CD: Pa nirḍi	*nir-Vṭ-	3674
LEARN	kal-	CD: Kl karp	*kal-	1297
SCRATCH, DRAW (LINE)	kīṟu	CD: Pa gīr 'line, mark'	*kīṟ-	1623
HUNTING	vēṭṭam	CD: Nk vēṭa	*[*vēṇ-ṭṭ-a-]*	5527
CUBIT (ELBOW/KNEE)	muẓam	ND: Mt muki 'cubit', múke 'knee'	*muẓ-k-	4990
WORK	paṇ	CD: Kl pannī	*[*paṇ]*	3884
TO PLAY	āṭu	CD: Nk āṛ-	*[*āṭ-u]*	0347

Notes

1 Glosses in bold type indicate cases with cognates in SD and ND, presumed to be the most reliable attestation for reconstruction to PD (see **8.23**). Wherever possible, the word given in the first column is from SD, usually SD1, and is Tamil if not noted otherwise. The word in the second column is from the most distant of the branches represented among the cognate words in the DEDR entry. Glosses are the same as the English gloss if not otherwise noted.

2 Reconstructed forms from Krishnamurti (2003 §1.2) are italicized and in brackets. The proto-meanings given here do not necessarily agree with those given by Krishnamurti. Reconstructions in Roman type are the author's.

3 The word for donkey, along with the two following words, suggests a familiarity with equids at this early period – though it may be noted that the two words for 'horse' have been questioned on formal linguistic grounds (see the following notes). Any of these words, if they existed in Proto-Dravidian, might have referred originally to the Asian half-ass, either the onager (*Equus hemionus*) or the khur (*E. hemionus khur*). According to Possehl, "the onager is native to the Greater Indus Region" (1999: 188), and its remains, among those of other wild ungulates, are found "in the zoological record of virtually every excavated [Indus age] site," in contexts suggesting that these animals were hunted for food (1999: 175). As Bökönyi notes, the khur is known "in the northwestern part of the Indian peninsula (Kutch and the Thar desert" (1997: 297). It is possible that the ass (*E. asinus*) was also among the equids known in this area (see Possehl 1999: 186).

(*Appendix notes A continued*)

Whether the true horse (*E. caballus*) was present at this period is questionable. There are claims for its presence at various Indus Valley sites, but there is no scholarly agreement on this. Meadow has stated that there are "no convincing reports" of actual horse remains "before the end of the second millennium B.C." (1987: 308). On the other hand, Meadow and Patel "do not doubt that horses came to South Asia at least by the end of the first quarter of the second millennium and *perhaps* even by the end of the third millennium," as indicated by figurine evidence and painted sherds dating to the beginning of the second millennium (1997: 309). The equid remains found in Mature Harappan levels at Surkotada in Kacch, discussed in Bökönyi (1997) and Meadow and Patel (1997), may have been from *E. caballus* (the "caballin theory") or from *E. hemionus* (the "half-ass theory").

Equid bones are attested at Hallur in Karnataka as early as 1600 BCE, becoming fairly common in the region by the early first millennium BCE (Bökönyi 1997: 299). Thus the presence of words referring to equids would perhaps not conflict with placing Proto-Dravidian in peninsular India any time after the beginning of the second millennium BCE. However, see **8.41** for a possible connection between Proto-Dravidian and the Southern Neolithic archaeological complex of the mid-third millennium BCE.

4 As a PD word this item is somewhat doubtful, as it occurs only in SD2 and CD languages which are more or less adjacent to each other. (The SD1 words given in DEDR 1711(a), Ta **kutirai** etc., are probably not connected. McAlpin (1981) has suggested a derivation from **kuti** 'jump' [DEDR 1705] – i.e. "jumper" – and a possible Elamite connection). Though similar in form to Indo-Aryan words like OIA **ghoṭa(ka)**, Hindi **ghoṛā**, this may well be a Dravidian word; Krishnamurti discusses a number of words with initial aspirates in CD languages (2003 §4.5.1.4, items 54–63).

5 Krishnamurti considers this etymology "doubtful" on phonological grounds (2003 §1.2.2.5), though to this writer it seems fairly straightforward. Given the attestation in Brahui and Tamil, the extreme geographic points where Dravidian languages are now found, the word would represent the earliest period of Dravidian.

6 Cf. Ch. 3 note 4 for early occurrences of this word.

7 The DEDR refers to similar words in Munda languages.

8 The use of this word to refer to the tiger (found in several languages in this entry) may be because of taboo.

9 Initial **c-** $\rightarrow \emptyset$ in SD1 (Krishnamurti 2003 §4.5.1.3).

10 This is another case where SD and CD agree on the meaning (cf. Kl sintā 'tamarind (tree)') while ND differs; in this case the agreement may be simply a result of retention.

11 As Weber has pointed out, ragi was an important food crop at the southern Harappan site of Rojdi (see **7.12**).

12 This reconstruction is put together from several different DEDR entries as described in **3.23A**, s. v. **vrīhi**. It is possible that the earliest Dravidian meaning was 'seed', which developed into the meaning 'rice' in CD and SD.

13 Cf. Ta vai 'straw, grass' DEDR 5553.

14 DEDR 5401 notes that the Malto word here may be a borrowing from IA.

15 Cf. Br nusxal 'grindstone' DEDR 3728 (xal 'stone').

16 In this entry the Tamil **erukku** means 'cut, hew, strike, etc.'; there is also a questionable Mt word **engde** 'to clear away weeds'.

17 In South Asia the English term ‘jungle’ often refers to any land which has not been brought under cultivation.

18 Note Ko **vē-gaṛ** ‘field for shifting cultivation’, Nk **vāyur**’highland field for cultivation’, suggesting possible earlier meanings.

19 Old Tamil **kōn̲** (S. Steever p.c. Oct. 1985). The meaning ‘palace’ in this entry can only be reconstructed to the PSD level.

20 Additional words from DEDR 5545 include Ta. **vēḷir** ‘a class of ancient chiefs in the Tamil country, the Cālukyas [early Tamil kings], petty chiefs’; **Ku belxā** ‘kingdom’, **belō** ‘queen of white-ants’. S. Steever comments: **vēḷ** itself may mean ‘spear’; **vēḷan̲** would be ‘chief’.

21 S. Steever writes: “The Tamil cognate of **K. qerme** ‘destroy’ could well be **kol/kon̲r̲-** ‘kill’ [DEDR 2132] instead of **cer̲u**.” (p.c. Oct. 1985). Both of these entries appear to involve the notion of killing human beings.

22 This word was provided by S. Steever (p.c. Oct. 1985). It is not listed in DEDR 5459.

23 S. Steever comments: “**uri-mai** is an abstract derivative from **uri-** ‘to own, be proper to’ ” (p.c. Oct. 1980).

24 S. Steever comments: “**ankāṭi** is probably a compound, **akam** ‘inner’ **kāṭu** where **kāṭu** means ‘building’ from the verb **kaṭṭu** ‘to tie, bind, thatch’ “ (p.c. Oct. 1985). [cf. also Tamil **kaṭṭaṭam/kaṭṭiṭam** ‘building’]

25 Kinship terms have been discussed by Krishnamurti (2003 §1.2.2.3) and are omitted here. See also Trautmann (1981) and Tyler (1968).

26 The Tamil and Kurukh glosses for this term are similar: Ta ‘man, husband, servant, labourer, adult’; Kur ‘an adult male person, husband, friend, servant, soldier’. The gloss ‘servant’ occurs for most of the languages represented in the entry. The basic meaning, however, appears to be ‘man’ (= adult male), the other meanings being extensions of this.

27 As a place name suffix, **-ūr** occurs in the four literary Dravidian languages (DEDR 752, George 1986).

28 **kuppam** is a common place-name suffix throughout South India (see George 1986). S. Steever writes: “The root **kup(p)-** seems to mean ‘gather, accumulate’ – the Tamil word **kupp-ai** means ‘garbage swept in a pile’. **kuppam** might mean an unplanned settlement” (p.c. Oct. 1985).

29 This is the basic word for ‘mouth’ in most SD languages. The meaning ‘edge of knife/sword’ is found in Ta Ma Ko Ka Tu Te Go (SD), Ga (CD), and Br (ND). It is possible that this is a case of homonymy in PD. This notion is reinforced by the presence of other words beginning with ***vā-** or ***va**, such as PD ***vaṭ-** ‘sharpen, sharpness’ DEDR 5349, PSD ***vāl** ‘sword’ (see Appendix B, G2), and Tamil **va(y)cci**, Tulu **bācci** ‘adze’ DEDR 5339.

30 Though this word is acceptable phonologically and semantically as a PD reconstruction, it is isolated. Another word for axle, borrowed from Indo-Aryan, may perhaps be reconstructed for PD (see **3.22B, 8.24**). Words for ‘wheel’ and ‘cart’ can only be reconstructed to the PSD level, though of course they may possibly be older: see App. B, G3, **8.24**.

31 In this expression, Ko **aḍ** means ‘at full length’ (DEDR 83).

32 The semantics here are unclear. The SD and CD languages agree on the meanings ‘headload’ and/or ‘to carry on the head’. The meaning ‘carry on the head’ also occurs in Kurux and Malto, along with the meaning ‘take upon oneself, take responsibility for’ – which also occurs (independently?) in Tamil. Brahui **kubēn** means ‘heavy’, agreeing with one of the meanings of the Tamil word. Perhaps the original meaning is ‘load’ or ‘carry, lug’. Note the possible parallelism with English **heave, heavy**. (Carrying heavy loads on the head is a pan-South Asian custom.)

Appendix B: vocabulary of Proto-South Dravidian

To be included in this list, an etymon must be represented in both SD1 (Tamil, Malayalam, Kota, Toda, Kannada, Kodagu, or Tulu) and SD2 (Telugu, Gondi, Konda, Pengo, Manda, Kui, or Kuwi). The most secure cases are those in which one or more non-literary languages are represented in both branches. In a number of cases, SD2 is only represented by Telugu; in these cases, when the Telugu cognate closely resembles the Tamil or the Kannada cognate, the etymology is less secure, since Telugu is known to have borrowed words from both Kannada and Tamil in the past. For these cases the gloss is given in ordinary type, whereas the more secure cases are glossed in bold type. Asterisked items are those which are attested in several SD1 languages, including non-literary languages, but not in SD2 apart from Telugu.

English gloss	*SD1 (default: Tamil)*	*SD2 (default: Telugu) {Comment}*	*Reconstruction*	*DEDR no.*
A. Fauna				
BUFFALO (FEMALE)	erumai	enumu	*[*erum-]*	0816
TIGER/LEOPARD	vēnkai	vẽ̄ngi	*vēn-kay	5521
CHAMELEON	Ka toṇḍe, toṇṇe, ḍoṇṇe	toṇḍa	*[*toṇṭ-]*	3501
SPIDER	Ko ko·j	Kd gōnzru	*kō(n)c-	2195
CHICKEN	kōḻi ‘gallinaceous fowl’	kōḍi	*[*kōḻ-i]*	2248
VULTURE	paruntu ‘kite’, perāntu ‘vulture’	pata-gadda ‘vulture, eagle’	*[*par-Vntu]*	3977
TURTLE/TORTOISE	(y)āmai ‘turtle, tortoise’	tāmēlu ‘id.’	*[*yām-ay, *cām-p-]*	5155
FROG	Ko kepak ‘frog, toad’	kappa ‘frog’	*[*kapp-a]*	1224
LIZARD	palli ‘gecko’	Kd balli ‘lizard’	*[*pall-i]*	3994
CRANE/STORK	vakkā ‘crane’	vakku ‘crane’	*vakk-	5206
*RAM	takar ‘sheep, ram, goat, male…’	tagaru ‘ram’	*[*tak-ar]*	3000
*CAT (DOMESTIC)	cīla “pussy” (calling)	cīli (name for pet cat)	*cīl-	2634
**FELIS* spp.	ciṟuttai ‘panther’	ciṟuta ‘leopard, cheetah’	*[*kiṯ-u-tt/mp-]*	2590
*SHARK	cuṟa(vu) ‘shark, swordfish’	coṟa ‘shark’	*[*coṯ-ac-]*	2710
B. Flora				
COPRA	kopparai	Kw kobri	*koppa-ray	2105
JACK (fruit)	cuḷai	Kw tōla	*coḷ-ay	2704
NEEM	vēmpu	Kw vēpa	*[*vē-mpu]*	5531

SOAPNUT	cīkkāy	Kd sīko	*[*cīk-kay]*	2607a
ARECA NUT	aṭai-kkai	aḍa	*[*aṭ-ay-kkāy]*	0088
JUJUBE	koṭṭai	goṭṭi	*koṭ-ṭay	2070
KINO TREE, *Pterocarpus marsupium*	vēnkai	Ki vēngesi	*vēn-kay	5520
AGASTI TREE, *Agasti grandiflora*	akatti	agise	*aka-c-tti	0005
PANDANUS, ***P. odoratissimus***	kaitai, kaital	gēdāgi	*kay-tay	2026
PANDANUS, ***P. odoratissimus***	mukari	mogali, (inscr.) movali	*mo-ka-ri/li	4890
CLEARING-NUT TREE, *Strychnos potatorum*	il(li), illam	cilla	*cil(1)-	2560
MASTWOOD	puṇṇai	ponna	*[*puṇṇ-ay]*	4343
WOOD-APPLE	veḷḷil	velāga	*[*veḷ-]*	5509
MUSTARD[1]	aiyavi	āva {see note}	*ay-a-	0921
ORANGE	kiccili (fruit)	kiccili, kittali (tree)	*kicc-il-	1512
CITRON LEMON, *Citrus medica*	mātaḷai	mādi	*māt-aḷ-	4808
ORANGE	īḻe	īḍe	*īẓ-e	0552
FLAX	Ka agase	agise, avise {cf. OIA ataśī}	*ak-V-ce	0003
OLEANDER, *Nerium odorum*	kaṇavīra	geṇṭena {cf. OIA karavīra (MBh), **3.23A5**}	*kaṇa-	1164
PALMYRA FRUIT (tender/unripe)	nunku	nungu	*nunk-u	3698
CUSCUS GRASS, *Andropogon muricatus*	viẓal	viḍavali {cf. OIA viraṇa (MBh), **3.32A5**}	*viẓ-al	5428

(*Appendix B continued*)

English gloss	*SD1 (default: Tamil)*	*SD2 (default: Telugu) {Comment}*	*Reconstruction*	*DEDR no.*
COIR (coconut fiber)	kayiṟu 'rope, coir'	cēru	*kaiṟu	1254
*COCONUT	teṉṉai, tenku (tree)	tenkāya (nut), tē-mrānu (tree)	*[*ten-kāy]*	3408
*FICUS (*F. glomerata, F. religiosa*)	atti	atti {cf. OIA aśvattha ← aśva- + att- ?}	*att-i	0144
*SANDALWOOD	cāntu	cātu	*cāntu	2448
C. Habitat				
FOREST	kāṭu	Kd kāṛ(u)	*[*kāṭu]*	1438
(SEA)SHELL	ippi 'pearl-oyster, shell'	cippa 'shell'	*cipp-	2535
MOUNTAIN[2]	kō	kōya 'Koya, a mountain tribe'	*[*kō-]*	2178
RIVER/STREAM	Ka toṟe	toṛe (in place names)	*toṯe	3370
FLOOD	veḷḷam	velli {← PD FLOOD/TORRENT, App. A, C}	*veḷḷ-	5503
LAKE/POND	Ka beñce	venca	*veñc-	5471
FOREST	paẓuvam	paḍuva	*paẓu-va-	4006
FOREST	kā(l), kāṉ(am)	kāna	*[*kāṉ-]*	1418
*OCEAN	kaṭal	kaḍali	*kaṭ-al	1118
*NORTH[3]	vaṭakku 'north(ward)'	vaḍāku-gōṇḍa 'Himalaya'	*[*vaṭ-akku]*	5218
D. Agriculture				
D1. Crop plants				
MILLET, *Panicum italicum*	ārike	āruka, aruka	*ār-V-k-	0379
FENUGREEK	mentiyam, mēntiyam	Go mētiŋ	*[*mēntt-i]*	5072
RICE/PADDY	campā 'superior kind of paddy'	sambāvu 'fine sort of rice'	*campā	2346
MILLET (species uncertain)[4]	varaku *Paspalum scrobiculatum*	varaga *Panicum miliaceum*	*var-ak-	5260
MILLET (species uncertain)	kampu	kambu	*kam-pu	1242
Dolichos catjang/ biflorus/sinensis	Ka alasandi	alacandalu	*ala-cant-	0242

DOLICHOS/BEAN	cikkaṭi 'field bean'	Kd cikuṛ 'kind of bean'	*cikk-Vṭ-	2496
CANNABIS	Ka gōgi	gō̃gu, gōnu	*kō(n)k-u	2183
PUMPKIN/GOURD	curai 'calabash climber'	sora 'bottle gourd'	*cur-ai	2690
*RAGI (*Eleusine coracana*)	irāki	rāgi {← PD FOOD, App. A, D1}	*[*ār/ar-ak]*	0812
*GREENS	kīrai 'greens, *Chenopodium album*'	kīra, kīre 'greens'	*kīrai	1617
*CARDAMOM	ēlam	ēla(ki)	*[*ēl-]*	0907
*INDIGO	aviri/avuri	aviri	*aviri	0269
D2. Irrigation				
TANK (RESERVOIR)	kuḷam 'tank, reservoir'	kol 'tank'	*[*kuḷ-am/-Vnc-]*	1828
BUND	Ka. uddi 'low bund'	uddi 'bund…'	*utt-i	0622
SLUICE	kalinku	kalũju	*kal-V-nk/nc-	1309
SLUICE	mataku	madugu	*[*mat-Vku]*	4688
*IRRIGATION CHANNEL	kāl(-vāy) 'irrigation'	kālava 'canal'	*kāl-av-	1480
*BANK/DAM	aṇai 'embankment, dam'	ana	*aṇ-ay	0122
D3. Other agricultural terms				
AGRICULTURAL LAND	paṇṇai	pan(n)asa 'agr. land given to Brahmins' (inscr.)	*[*paṇ-]*	3891
PLOUGHSHARE	kāṟu	kaṟṟu	*[*kāṯ]*	1505
PLOUGH HANDLE	mēḻi	mēḍi	*[*mēḻ-i]*	5097
SEEDLINGS	nāṟu	nāṟu	*[*ñāṯ-u]*	2919
GRAIN MEASURE	puṭṭi	puṭṭi	*[*puṭṭi]*	4262
TO CARD (COTTON)	Ma ekkuka	ēku	*ekk-u-	0765
*PADDYFIELD	kaḻani	gaṇḍamu {ND 'sediment'}	*[*kaḻ-]*	1355
E. Political and economic relations				
TAX/TRIBUTE	ari 'tax, duty'	ari 'tax,tribute' {see **8.20A**}	*[*ar-i]*	0216
LORD/MASTER	iṟai 'chief, illustrious person'	eṟa 'lord' (inscr.)	*[*eṯ-ay-anṯu]*	0527
CHIEF	maṉ 'king'[5]	manniya 'lord, chief'	*maṉ-	4774
CHIEF	mēṭṭi 'chief, headman'	mēṭ(ar)i 'chief, head, lord' {*mēṭṭi←PD *mēl-ti 'over, above' DEDR5086}		5091

(*Appendix B continued*)

Appendix B Continued

English gloss	*SD1 (default: Tamil)*	*SD2 (default: Telugu) {Comment}*	*Reconstruction*	*DEDR no.*
*TO RULE	āḷ 'rule, govern, manage'	ēḷu 'rule' (inscr.), ēlika 'ruler'	*[*yāḷ]*	5157
ARMY/BATTLE	paṭai 'army'	paḍava 'battle'	*paṭ-ay/av-	3860
COIN/CASH	kācu 'gold, coin, cash'	kāsu 'cash, coin'	*kāc-u	1431
COMMODITY/WARE	carakku	Kw harku	*car-a-kk-	2353
TRIBUTE	kappam	kappam	*[*kapp-am]*	1218
BATTLE (ARRAY)	aṇi 'array of an army'	aṇi 'battle, war' {? ← OIA anīya}	*[*aṇi]*	0117
WAR/BATTLE[6]	pōr (poru 'to fight')	pōru 'to fight; war' {CD: Ga pōr- 'abuse, scold'}	*[*pōr]*	4540
BATTLE(FIELD)	kaḷam, kaḷan	kalanu 'war, battle, combat' {← PD THRESHING-FLOOR, App. A, D3}	*[*kaḷ-an]*	1376
BATTLE	muṉai 'battle(field), etc.'	mona 'battle array' {orig. mg.: 'anger, angry'}	*[*muṉ-ay]*	5021
CUSTOMS/TOLL	meṭṭu 'customhouse'	meṭṭu 'toll-gate'	*meṭṭ-u	5059
TERRITORY[7]	nāṭu 'country, locality, place, etc.'	nā(ṇ)ḍu 'country'	*[*nāṭu]*	3638
DIVISION OF TERRITORY	maṭappam 'town in agr. tract, chief town…'	maḍambamu (inscr.) 'geog. division'	*maṭ-a(m)p-	4646
WAGER[8]	pantayam	pandamu	*pant-	3921
UNPAID LABOR	amanci	amanji	*aman-ci	0159
*KING	aracaṉ 'king'	arusu {← IA rājan?}[9]	*arac-aṉ	0201
*MANAGER	maṇiyam 'office of village headman'	maṇiyamu 'office of temple manager'	*maṇi-ya-	4674
*ARMY	taṇṭu	daṇḍu {orig. mg. 'crowd, herd'?}	taṇṭ-u	3055
*PRICE/TO SELL	vil-(viḻ-)/vilai	vil(ucu)/vela	*[*vil-]*	5421
*COLLECT (MONEY/DEBT)	taṇṭu	daṇṭu	*taṇṭ-u	3054
F. Social relations				
F1. Caste/occupation				
CASTE (SPECIFIC)[10]	kuṟavaṉ 'man of a caste of fowlers… and fortune tellers'	kuruma 'a caste of shepherds'	*koṟa-	1844

BASKETMAKER (CASTE)	ka mēda	mēdara	*mēt-ar-	5092
WEAVER (CASTE)	cāliyan̲	sāle {← PD ‘spider’}	*cāl-	2475
TODDYTAPPER	īẓavan ‘toddy-drawer’s caste’	īṇḍra	*[*īẓ-a-vant]*	0549
WEAVER	cēṇṭiravan̲	jēṇḍra	*cēnt-ir-	2809
FARMER	veḷḷāḷan̲ ‘agricultural caste’	velama {cf. PD CHIEF, App. A, E)	*vel(ḷ)-āḷ-	5507
OUTCASTE	mālar ‘outcaste’	māla ‘pariyah’	*māl-	4824
SERVANT/SLAVE	toẓukkan̲, tottan̲ ‘slave’, toẓuvar ‘servants’	tottu ‘female servant/slave, wench,…’	*toẓ-u-kk/tt-	3524
*SCAVENGER/PARIAH	tōṭṭi ‘menial servant, scavenger’	tōṭi ‘pariah’ {CD: Pa tōṭa ‘laborer’}	*tōṭṭ-	3546
F2. Other				
GUEST/VISITOR[11]	viruntu	vindu	*vir-unt-	5415
G. Material culture and technology				
G1. Urban habitat				
TOWN/CITY	poẓal ‘town, city’	prōlu	*poẓ-al	4555
VILLAGE	ūr ‘village, town, city’	ūr ‘village’ {← PD ‘house’, App. A, G1}	*[*ūr]*	0752
HAMLET/VILLAGE	paḷḷi ‘hamlet, herdsman’s village,… ‘	palli ‘village, hut’	*[*paḷḷi]*	4018
VILLAGE/TOWN	pūṇṭi ‘town, village, district’	-pū̃ḍi (suffix for small village or hamlet name)	*pūṇṭ-i	4362
ABODE	paṭṭi ‘cowstall, sheepfold, hamlet, village’	paṭṭu ‘dwelling place’ {← *paṭu ‘lie down’?}	*paṭ-ṭ-	3868[12]
HAMLET[13]	cēri ‘town, street, quarter’	gēri ‘street, passage’	*kēr-i	2007
HOUSE/ABODE[14]	nakar ‘house, abode, temple, palace, city’	nagaru ‘palace, temple’	*nak-ar	3568
HABITATION[15]	pākkam ‘village, (seaside) town’	pāka ‘hut’	*pākk-	4047
GATE OF FORT/CITY	kavun̲i ‘gate of fort/city’	gavanu ‘gate of fort’	*kav-Vn̲-	1339
FORT/CASTLE	kōṭṭai ‘fort/castle’[16]	kōṭa {← PD WALL, App. A, G1}	[*kōṭṭ-ay]	2207a

(*Appendix B continued*)

English gloss	*SD1 (default: Tamil)*	*SD2 (default: Telugu) {Comment}*	*Reconstruction*	*DEDR no.*
PALACE (also 'temple')	kōvil	kōyila, kōvela {= *kō 'king' + *il 'house'}	*[*kōy-il]*	2177
WALL, COPING	matil 'fort wall; coping'	maduru 'wall coping'	*mat-V-l-	4692
PLATFORM/VERANDAH	tiṇṇai	tinniya	*tiṇṇai	3227
TILE (ROOFING)[17]	Ka hencu	pencu, penku	*[*pen-kk-]*	4385
PRISON/STOREHOUSE	kiṭanku	giḍḍangi	*kiṭ-a-nk-	1525
MOAT	akaẓi	agaḍta 'ditch, moat, trench' {ND, CD 'dig, ditch'}	*[*akaẓ-tt-ay]*	0011
LANE	Ka. ōṇi	ōni-kaṭṭu 'dark alley, narrow pass between hills' (earlier mg.?)	*ōn-i	1046
LANE	Ka gondi, gondu 'alley, lane,…'	gondi 'corner, lane'	*kon-t-V	2100
G2. Metals and metalwork				
TIN	takaram	tagaramu, Kw ṭagromi	*tak-a-r-	3001
ANVIL	aṭai-kal	ḍa-kali	*aṭa + kal-{= 'support-stone'}	0086
SMITHY/WORKSHOP	paṭṭaṭai 'anvil, smithy, forge'	paṭṭaḍa 'workshop'	*paṭṭ-aṭ-ay	3865
IRONWORK	kol 'ironwork(er)'	Kw kolhali 'to forge'	*[*kol-]*	2133
SMITHY	Ka kolime	Go kolmi	*kol-i-m-	2133
SWORD	vāḷ	vālu	*[*vāl]*	5376
SWORD HILT	parincu	parũju	*par-V-nc-u	3969
WEAPON	kaidu, keydu 'weapon' (Tu kaidu 'sword')	kaiduvu	*kay-tu	2027
ARMOR	kattaṇam	kattaḷamu	*katta-ṇ/ḷ-	1203
HASP/LOCK	cilaka	cilaka	*cila-kk-	2561
HAMMER	cutti	Go sutte	*cutti	2668
SLEDGEHAMMER	cammaṭṭi	sammeṭa	*cam-ma-ṭṭi	2349
REFINED GOLD	aparanci	aparanji	*apa-ra-nc-i	0152
INGOT	pāḷam	pāḷā	*pāḷ-	4114

CHAIN	kolucu	golusu	*kol-u-cu	2134
*WIRE	kampi ‘wire’	kambi ‘wire’, kammi ‘line on cloth’[18]	*kamp-i	1241
G3. Transport				
WHEEL	kāl	kalu {← PD *kāl ‘leg’, DEDR1479}	*kāl-	1483
WHEEL[19]	uruḷ(i)	Kd uṇḍa {= ‘to roll, something that rolls’}	*ur-uḷ-(nt-)	0664
STERN OF VESSEL	amaram	amaramu	*am-ar-am	0163
BOAT	paṭaku	paḍava	*[*paṭ-Vku]*	3838
*SHIP	kappal ‘ship’	kappali	*[*kapp-al]*	1219
*SHIP	kalam ‘vessel, plate, … ship’	kalamu ‘ship’, kalamari ‘sailor’ {← PD POT, App. A, G4}	*[*kal-am]*	1305
*CART/CHARIOT	tēr ‘car, chariot’	tēru	*[*tēr]*	3459
G4. Other				
PEARL	muttu	muttiyamu	*mutt-u	4959
PEARL FISHERY	calāpam	salāpamu	*cal-āp-	2369
UMBRELLA	kuṭai	Kw guṛgu	*[*koṭ-ay]*	1663
THREAD	nūl ‘cotton thread’	nūlu ‘id.’ {ND ‘to twist’}	*[*nūl]*	3726
THREAD	iẓai	Ki ēs- ‘spin thread’	*[*eẓ-]*	0506
COTTON	parutti	p(r)atti	*[*par-utti]*	3976
ROPE	Ma. kayaṟu ‘rope’, cēri ‘coconut husk/fibre’	cēru ‘string, cord’	*kayar-	1254
POT	kuṭam ‘waterpot’	kuḍa (inscr.) ‘small earthen vessel’	*kuṭ-a	1651
*POT	pāṇai ‘large earthen pot’[20]	bāṇa	*[*vān-ay]*	4124
MORTAR RIM	kuntāni	kund(ene)	*kunt-	1726
SPEAR/JAVELIN	Ma cilukku	cillakōla	*cil(l)-a-k-	2568
ARROW	kaṇai	Go kaṇī	*kaṇ	1166
BLACK	kāẓ	Md kaṛindi	*kāẓ	1494
BODICE	(i)ravikkai	ṟavika	*ṟavi-kk-ay	5163
NECKLACE	uttaṇṭa-maṇi ‘women’s gold necklace’	uttaṇḍamulu (plur.) ‘id.’	*utt-a-ṇṭ-	0618
SALTPAN	aḷam	upp-aḷama (uppu ‘salt’, cf. SALT in App. A, G4)	*aḷam	0299

(*Appendix B continued*)

Appendix B Continued

English gloss	*SD1 (default: Tamil)*	*SD2 (default: Telugu) {Comment}*	*Reconstruction*	*DEDR no.*
PITFALL (FOR ELEPHANT)	Ka ōda, vāda	ōdamu	*ōt-a-	1048
CUSHION (?)	tiṇṭi 'cushion, brick support'	diṇḍu 'pillow, bundle'	*tiṇṭi	3224
*GARLAND	mālai	māla	*māl-ay	4827
*HORN (MUSICAL)	kompu	kommu {CD 'animal horn'}	*komp-u	2115
H. Religion				
PRIEST/TEMPLE SERVANT	Ka tammaḍi 'attendant on idol'	tammaḍi 'priest of a Siva temple'	*tamm-a-ṭi	3085
DEMON/GOD	pēy 'devil, goblin, fiend…'	Go pēn 'god' {= PD 'possession', App. A, H}	[*pēy, *pēn]	4438
SACRIFICE/WORSHIP	vēḷ-/viḷai	vēlucu/vēlpudu (vēlpu 'god')	[**veH-ḷ-]*	5544
WORSHIP/HOMAGE	toẓu 'worship, adore, pay homage to'	Go doṛī 'bow'	*toẓ-	3525
*TEMPLE	kōvil	kōyil, kōvela	[**kōy-il]*	2177

Notes

1 Note also Ka **sāsavi**, Te **sāsavulu** (DEDR 0921); possibly ← OIA **sarṣapa** (see Burrow 1955: 379).

2 This word is solidly represented in SD2, but in SD1 occurs only in Ta.

3 Other compass directions: the word for 'south' can only be reconstructed to the PSD1 level: PSD1 ***teṉ** 'south' (DEDR 1649). Words for 'east' and 'west' are found in individual languages, and are derived from other meanings: Go **siṛāyiŋ** 'east' (**siṛ** 'below') ← PSD ***cir-V-tt-** 'the low area' (DEDR 2584); Ta **kiẓ-akku** 'east(ward)' ← PD **kīẓ-** 'down, below' (the meaning 'east' occurs only in Ta Ma Ko Kod); Ta **mēṟku** 'west(ward)' ← PD ***mē-**'up, above' ('west' in Ta Ma Ko only); Ta **kuṭaku** 'Coorg, west' ← PSD1 ***koṭ-**, a word referring to an ethnic group, the Coorgs – perhaps also to the Kudux (DEDR 1649). Note that Kannada **baḍa(ga)** 'north' is also the name of the Badaga tribal group.

4 See the discussion of the shifting meanings among different millets in **8.4**.

5 Probably derived from the verb **maṉ** 'be located': cf. **maṉ-ai** 'house', **maṉ-ram** 'council', **maṉ-nan** 'prince' (i.e. 'one who is well placed/located'), etc. (S. Steever p.c. Oct. 1985).

6 This is a PD word meaning ‘to dispute, argue; dispute, argument’. Only Ta Ma Ka Tu Te have the meaning ‘battle’, suggesting that this is probably a word belonging to the literary period. The next two words also have the meaning ‘battle’ or ‘battlefield’ only in the literary languages.

7 The non-literary languages in both SD1 and SD2 have the meaning ‘village, inhabited place (as opp. to jungle)’, which is probably the earlier meaning; ‘country, territory’ is found only in the literary languages.

8 The older meaning here is probably ‘race, contest’, which is preserved in Ko (SD1) and Kw (SD2). This may be connected to an OIA word, **panita** ‘wager’ (see DEDR 3921).

9 Burrow and Emeneau list this word in the main body of the DEDR, rather than in the Appendix, with the notation “< Indo-Aryan.” It occurs in all the SD1 languages, with meanings which include ‘king, sovereign, kingdom, government, palace, court, royalty, constitution’, etc., implying a longish period of development within SD1. There is a single Telugu form, **arusu** ‘king’, and no other SD2 cognates.

10 The origin of these terms is indicated by the other parts of this entry: Ta kuṟam ‘Kurava tribe’, Te koṟava ‘name of a tribe of mountaineers’, Go kurmal ‘shepherd’. Thus it is not totally clear that the PSD form referred to caste.

11 The original meaning may be ‘stranger, person from another village’.

12 This is also a common place-name ending in South India (George 1986). For its possible use in Maharashtra, see **9.23**.

13 ‘Settlement’ (i.e. village or hamlet) is probably the original meaning of this term. Note that other words have undergone the meaning shift ‘settlement’ → ‘quarter of a town’→ ‘street’ (see above).

14 Perhaps initially ‘house, temple’.

15 Initially ‘hut (of low-caste person?)’.

16 This is also a common ending of place names in the four southern states (see George 1986).

17 Perhaps initially ‘potsherd’.

18 Te **kammi** ‘line on cloth’ may show the earlier mg. (also found in Ta); **kambi** ← Ta **kampi** [kambi]?

19 It is questionable whether the mg. ‘wheel’ can be reconstructed to PSD, though it occurs in both SD1 and SD2, because of the likelihood of borrowing or independent derivation of this mg. from the verbal mg. ‘to roll’ and other mgs. like ‘ball, rolling object’. In this case, this mg. may have originated in SD2, as it occurs in Te, Kd, and Kw (specifically ‘wooden wheel’).

20 In the non-lit. SD1 lgs Ko and Kg, this word has the mg. ‘grain measure’.

Appendix C: selected reconstructions in PSD1 (see 8.43)

A. Sources of OIA loanwords

The following words have been suggested as sources of OIA loanwords (see **3.22A**). Although these proposals have not been accepted by all scholars, the PSD1 words may nevertheless indicate early contact between PSD1 and OIA, possibly involving a third language as ultimate source in some cases (see **3.23A**).[1]

PSD1 cognates	*PSD1 reconstruction*	*OIA form*	*DEDR no.*
Ka aṇile, Tu aṇile 'inknut tree'[2]	*aṇile (Proto-Tamil)	nīla (RV) 'blue', **3.23A2**	0119
aṉal, Ka analu 'fire, heat'[3]	*aṉal	anala (lex.), Pali anala 'fire', **3.23A4**	0327
ulakkai, Ka olake 'pestle'	*[*ul-akk-V]*	ulūkhala (RV) 'mortar', **3.23A2**	0672
erukku, Ka ekke *Calatropis gigantea*	*eru-kk-	arka, **3.23A4**	0814
iñci 'ginger'	*[*(c)iñk-i]* (Proto-Tamil)	śṛṁgavera (Suśr), Pali singivera, **3.22A4**	0429
Ta, Ka tippali 'long pepper'	*tipp-a-li	pippali (RV), **3.22A1** s.v. phala-	3228
nāram(-kāy), Tu nārengi 'orange'	*nār-am(-kāy)	nāranga	3653
miḷaku, Ka meṇasu/moẓasu 'black pepper'	*[*miḷ-Vku]*	marīca (Cl) 'peppercorn'	4867
vā(y)cci, Ka bāci 'adze'	*vāy-cci	vāśī (RV), **3.22A** note 4	5339
vāẓai, Ka bāẓe 'banana, plantain'[4]	*[*vāẓ-a]*	{cf. OIA vārabusā *Musa sapientum*}	5373

Words with questionable Telugu cognates:
***atti** 'fig' App. B, B: OIA aśvattha,[5] ***ēl-** 'cardamom' App. B, B: OIA elā (Suśr), ***mālay** 'garland' App. B, G4: OIA māla App. B, A4, ***vaḷḷī** 'creeper': OIA **vallī** 'creeper' (Mn) **2.23A5**, ***kaṇa-vīra** 'oleander' **App. B, B**: OIA karavīra (MBh), **3.23A5**, ***var-ak-** 'millet' App. B, D1: OIA varuka 'an inferior grain'.

B. Sources of words in languages outside of South Asia

The following PSD1 words are among the many Dravidian words (see **7.24**) which have found their way into Arabic and European languages, mainly as a result of trade:

PSDI cognates	*PSD1 reconstruction*		*DEDR no.*
verr-ilai 'betel leaf' (ilai 'leaf')	*vet̲t̲- (proto-Tamil)	{Port. **betel**, etc.}	5515
cakkai, Ka jaka 'jack' (tree/fruit)	*cak-kay	{Port. **jaca**, English **jack**}	2275
Ta, Ka eḷ(ḷu) *Sesamum indicum*	*eḷ	{Akkadian **ellu**, **7.12B9**}	0854
kayir̲u, Kd cēri 'coir'	*kay-ir- (proto-Tamil)[6]	{Port. **cairo**, E **coir**}	1254

Note also (from (A)): ***iñci*** 'ginger' (Gk **zinziber**, Fr **gingembre**, etc.), ***aṇile** 'inknut tree' (E **anil, aniline**), ***nār-am** 'orange' (Farsi **nārang**, Arabic **nāranj**, Fr **orange**, etc.).

Notes

1 See also **8.24** for PSD borrowings from OIA.

2 Another word for this tree is PD ***kaṭu-kkāy***, see *Terminalia chebula* in **App. A, B**.

3 There are many words referring to 'fire' and 'heat' in the DEDR: examples are PD ***kiccu** 'fire' DEDR 1514, PD ***cuṭ-** 'be hot, burn' DEDR 2654.

4 This is the most widely distributed word for 'banana' or 'plantain' in the DEDR. Other words are: ***ar-Vṇṭi*** (Te and Go only) DEDR 205, DEDR 754 (Pa and Ga only), DEDR 3181 (Kd, Ki, Kw only).

5 Other words for *Ficus* spp. include PD ***tōnk-**, ***mēḍi** (App. A, B).

6 This word occurs in PSD with the meaning 'string, cord' (**App. B, G4**). The meaning 'coir' is restricted to Proto-Tamil, suggesting that the English **coir** came from the extreme south of India; see **7.24**.

Notes

1 Tamil is attested in cave inscriptions from the second century BCE. The earliest known grammar of Tamil, the Tolkappiyam, is believed to be from a late century BCE; it presupposes the existence of a "large body of literature before it" according to Krishnamurti (2003 §1.4.1, v. section 1. TAMIL). Other scholars consider the earliest Tamil literature to belong to the early centuries of the CE (see **8.42**). Telugu names appear in Prakrit inscriptions from the second century CE onwards, and the first inscription in Telugu is from 575 CE. Kannada appears in inscriptions from about 450 CE, Telugu in 575 CE, and Malayalam in the ninth century CE (op.cit. s.v.2. MALAYALAM, 3. KANNADA, 4. TELUGU).

2 PD *pāṭu 'recite, sing' is not glossed as 'read' in DEDR 4065.

3 Compare English **write** ← OE **wrītan** 'cut, scratch'; cf. German **reissen** 'to tear'.

4 The word ***pār** 'brahman' is not totally reliable as a PD reconstruction, since it is attested in only one item outside of South Dravidian: Nk. (Ch.) **pār** brahman, fem. **pāriya**, which might well be a borrowing from one of the literary languages: cf. Ka. **pārva, pārba, pāruva, hāruva, hārva** brahman; fem. **pārviti; pārvike, pārbike** brahmanism; Te. **pāruḍu, pāruvaḍu** id.; fem. **pāruṭa, pāruṭakka**. In any case, as an etymon attested in SD and CD only, it is not guaranteed to belong to the earliest period of PD. (The DEDR entry for this word suggests a connection with the verb 'see'.)

5 See also Ta **paṟaiyaṉ**, Ma **paṟayan** 'drummer; member of Pariah caste' ← PSD ***paṯ-ay** 'drum; grain measure' (DEDR 4032).

6 This word occurs in two forms: PSD ***ari** 'tax, duty, unjust payment, loan of grain, purchase on credit, debt, etc.' [DEDR 216, Ta Ka Te only] and PSD ***vari** 'tax, duty, contribution' (to government or to another group, e.g. tax paid by Badagas to Kotas) [DEDR 5266, Ta Ma Ko Te only]. It seems very probable that this word is related to another word which shows the same variation in form: PSD ***ari** 'rice, paddy'[DEDR 215], PSD ***vari** 'paddy', possibly originally meaning 'seed' (see RICE in Appendix A, D1). Payment in grain is still practiced in South India, for example, for payments of day-wages to agricultural laborers, especially for harvesting (Mencher 1978: 209–10). Thurston notes: "From the earliest times the Todas have received from the cultivating Badagas an offering or tribute, called gudu or basket of grain, partly in compensation for the land taken up by the latter for cultivation, and so rendered unfit for grazing purposes" (Thurston 1909: I. 137–8). Thurston also mentions an annual Kota ceremony in which Kota priests offer grain to Todas in return for gifts of ghee and blessings on their buffaloes and their crops (op.cit. Vol. IV, p. 19). Note also PSD ***vār-(ak-)am** 'land-rent, share of crop' ← PD ***vār-V** 'debt' [DEDR 5359]. In describing the relationships among the Nilgiri tribal groups, Hockings notes: "Each of the four main groups – Badagas, Kotas, Todas and Kurumbas – was linked with each of the others through patrilineally descending partnerships. The men involved in these supplied each other with goods and services supplementing their own economic production and also symbolizing their social relationship through ritualized interaction" (Hockings 1980: 131). These exchanges took place at certain festivals, and during rituals connected with marriage, death, etc. For example, "The Kotas were the specialists in leather-work, carpentry, pottery, metal-working, thatching and ceremonial music. Their participation in Toda and Badaga funerals was essential" (ibid.).

7 This word may be derived from PD ***eṯ-** 'transcend, be pre-eminent' (DEDR 515).

8 The meanings of this word in non-literary Kota and Toda refer to possession of diviners or shamans, suggesting that the word goes back to religious observances which preceded the use of carts to carry idols in procession.

9 Krishnamurti discusses additional shared innovations within Central and South Dravidian which are not represented in Figure 8.1, as they do not bear on the discussion here.

10 In two items (3 percent of the less secure cases), the semantics of the words suggest origin in Telugu or another SD2 language. (See the relevant endnotes.)

11 DEDR 5260 is glossed variously as *Paspalum scrobiculatum, P. frumentaceum*, and *Panicum miliaceum*; DEDR 1242 = bulrush millet, Italian millet, *Holcus spicatus, Pennisetum typhoideum, Panicum spicatum*; DEDR 379 = *Panicum italicum, Paspalum scrobiculatum, P. frumentaceum, Setaria italica*.

12 It may be relevant to point out that in the TRANSPORT category, only one word for ‘wheel’ is etymologically secure, while the remainder, including all the terms relating to ships and boats, belong to the less secure categories. (One of the words for ‘wheel’ appears to be of SD2 origin.)

13 In this case, it would also be possible to consider those OIA borrowings in Dravidian which appear only in SD1 languages (or with questionable Telugu cognates) as relevant evidence for OIA-Dravidian contact during the mid-to-late second millennium BCE. See, for example, the following items from the Appendix to DEDR (Burrow and Emeneau 1984: 509–14):

 26. Ta **kīl** ‘hinge, nail, pin, spike, wedge, etc.’ with cognates in Ma Ko To Ka Kg Tu Te: OIA **kīla** ‘stake, pin, etc.’ (P. 510);
 33. Ta **kōṇi** ‘sack’ (→E. gunny [bag]), with cognates in Ma Ko Ka Tu Te: OIA **goṇī**) p. 511);
 53(a) Ta. **maittu<u>n</u>a<u>n</u>** ‘brother-in-law, etc.’ with cognates in Ma Ko To Ka Kg Tu: OIA **maithuna** (p. 513).

14 See Chapter 9 for a discussion of the earlier Dravidian-speaking population in Maharashtra, especially along the coast.

15 The finding that skeletons of the Southern Neolithic archaeological complex show “uniform features…shared over the larger Deccan peninsula, the region south of the Narmada” which “continued into the Iron Age” (Fuller 2001a) is perhaps an argument against the foreign provenance of the Dravidian languages (see **8.41**).

16 Compared with the Indus Valley and other regions in northern South Asia, South India has been relatively neglected by archaeologists. Much of the excavation work that is reported in the literature goes back to the 1960s, but there are new studies currently underway which give promise of shedding new light on this region.

17 “This much however can be said that from their rich repertoire of pottery, the people cannot be regarded as primitive” (Sankalia 1974: 526).

18 These animals appear in rock-paintings and etchings, and even if their traces have not been found, one can argue that people who were capable of producing a realistic representation of an animal probably would have had a word for it in their language.

19 The existence of some “tuberous food” is inferred from the presence of parenchyma fragments (as yet unidentified as to species) at numerous sites (Fuller *et al*. 2001: 174–5).

20 Reported from Watgal (Raichur Dt, Karnataka), v. Fuller 2001a: 199; Devaraj *et al*. 1995).

21 “When considered on a general regional scale it can be seen that groups of permanent village sites and ashmound encampments form geographical clusters that might represent networks of agricultural villages with associated hinterlands of pastoral transhumance” (Fuller *et al*. 2001: 180).

22 Wheat is not a significant crop in South India nowadays. The words for wheat in the Dravidian languages are all borrowed from Indo-Aryan (OIA **godhūma**, Bangla **gohu**, Marathi **gahu**, etc.: cf. Kannada **gōdi**, Brahui **<u>kh</u>ōlum** (DBIA 123). These Dravidian words may have been early borrowings, since they show signs of the OIA **d(h)** which was lost already in MIA, and thus might be considered to belong to PD.

23 Fuller *et al*. (2001: 174, 180) note that remains of *Oryza* spp. are found in contexts which suggest the possibility that they were introduced from outside the area in the

later phases of the Southern Neolithic. Note that there is also a possibilty of *Oryza* having a separate origin in southwestern India **(7.12B8)**.

24 Thurston describes various ceremonies in which *Setaria italica* plays a role: for example, an annual agricultural ceremony performed by some Badaga groups known as **Devvē habba** or **tenai** (= *S. italica*) in which sheaves of the plant are used (Thurston 1909: Volume 1, p. 95), as well as the practice of throwing handfuls of the grain on the place where a body has been cremated (op. cit. 119).

25 Here again, as in the case of ragi (see RAGI in App. A, D1), a word for a cereal appears to have been derived from a word for food: cf. PD **tin̲-** 'eat' DEDR 3263(a).

26 These terms may relate to later developments, or might possibly have originally referred to different levels of terraces on the hillsides where the Southern Neolithic dwellings were constructed, and the vertical pathways connecting them.

27 "Since the Southern Neolithic province was mapped by Paddayya (1968, 1973)…the peninsular Neolithic has been extended into the major part of Andhra Pradesh (Middle and Lower reaches of the Godavari and Krishna basins) excluding the uplands of the Eastern Ghats, the whole of the east and southeast coast, mainland Tamilnadu excluding the Nilgiri and Western Ghats plateaus in Karnataka. Meanwhile the west coast, excepting the Goa region…still draws a blank and remains largely unexplored" (Fuller 2001a: 174).

28 The earliest canals identified in South India are from the second or third century BCE (Srinivasan 1991).

29 Coromandel, referring to the southeast coast of India, is an anglicization of Tamil **cōḷāmaṇṭaḷam**, meaning 'Chola country'.

30 The name Damirica is presumably derived from 'Tamil', possibly from **tamiẓakam**, an old word for the Tamil country.

31 The Greek **z** *ζ* seems to be the regular equivalent of the South Dravidian **c**, as seen in other words such as Gk. **óryza óρυζα**; 'rice' (←PD ***varici)** and **Muzíris Μυζιρις** for **Muciri**, a west-coast port (see Southworth 1979b). The Greek word seems to have come directly from the PSD1 form, indicating that the product was probably carried by sea directly to the west without passing through North Indian intermediaries.

32 This appears to be a synonym for **man̲ram** 'assembly hall', a PSD1 word whose original meaning probably is something like 'village gathering place', to judge by its meanings in the non-literary languages (DEDR 4777).

33 It is possible that PSD1 and PSD2 were, at least for some period of time, distinct subgroups, because of the large number of unshared innovations which occurred in each group, though they eventually reestablished contact. This argument is supported by the nature of the linguistic boundary between Telugu and its neighboring literary languages, Tamil and Kannada. This is a *bilingual* boundary, characterized by border bilingualism (like the boundary between Marathi and Kannada, or between French and German in Europe), rather than a *transitional* boundary, which perpetuates an older dialect continuum overlaid by distinct official languages (such as the boundary between Tamil and Malayalam, or between French and Italian). In this connection, it is of interest that the Telugu–Kannada boundary, which is part of the SD2–SD1 boundary, more or less coincides with the eastern edge of the Southern Neolithic core area.

34 In fact, it is conceivable that the "towns" of PSD were settlements of a type that would be considered large villages nowadays. In this connection, it is possible to imagine that the different quarters of the Chola port of Kāviripūmpattinam or Puhār described in the Sangam literature (see above) may have originally been adjacent villages; a number of modern cities, in South Asia and in other parts of the world, have formed around such village groupings. As noted in **9.23**, the semantics of the Tamil word **paṭṭiṇam** 'city, port city' (the suffix found in Kāviripūmpaṭṭinam) and the

related words **paṭṭi** ‘hamlet’, **paṭṭam** ‘sleeping place for animals’ (DEDR 3869), along with **pāṭi** ‘hamlet, village, town, city’, **pāṭam** ‘street of herdsmen’ (DEDR 4064), seem to recapitulate the entire history of human habitation, beginning with the temporary camps of nomadic or semi-nomadic cattle keepers, and including the transition from cattle shelters to human settlements, as well as the later incorporation of pastoral and agricultural hamlets into larger village complexes.

35 It would probably be useful to examine Paddayya’s (1973) regional divisions of the Southern Neolithic to see if there is any possible correlation with the linguistic divisions within South Dravidian.

9

MAHARASHTRIAN PLACE NAMES AND THE QUESTION OF A DRAVIDIAN SUBSTRATUM

Contents

9.1. Introduction

In any long-settled area, the names of places embody a great deal of the history of the inhabitants – past inhabitants as well as present ones. In South Asia, where even small villages have often been found to have histories measured in millennia, and where ancient records abound, this statement is more true than in many other parts of the world. And yet this great historical resource has not been adequately exploited. This chapter will examine some Maharashtrian place names, following up a suggestion made by Raymond and Bridget Allchin in the following passage from *The Rise of Civilization in India and Pakistan*:

> It has been suggested that the placenames of Maharashtra show a substratum of Dravidian elements, and these we may expect to relate to an earlier culture phase, such as that represented by the Malwa ware or

> the as yet little known pre-Malwa Neolithic phase which was akin to that of the south. We may therefore postulate that the original population of agricultural settlers was Dravidian speaking, and that the changes associated with the Jorwe period coincided with the arrival in the area of immigrants from the north speaking an Indo-Aryan language. This language must have been the ancestor of modern Marathi.
>
> (B. Allchin and R. Allchin, *The Rise of Civilization in India and Pakistan*, Cambridge: Cambridge University Press, 1982, p. 352)

The suggestion of a Dravidian substratum in Maharashtra is of interest for several reasons. I have suggested elsewhere (Southworth 1974) that a number of linguistic elements in Marathi point to the probability of such a substratum.[1] If it can be shown that Maharashtrian place names contain a significant Dravidian element, this would provide further support for the hypothesis of an earlier Dravidian-speaking population in this region, and might also make it possible to locate this population more precisely in place and time. Another major reason for interest in this subject is that toponymic studies, in India as elsewhere, have often been carried out in isolation from other types of linguistic and historical studies. Thus in the context of the present work, an investigation of this particular question will provide an opportunity to integrate the evidence of place names with other types of evidence used in linguistic archaeology.

What must we look for in the place names of Maharashtra in order to test the hypothesis of a Dravidian substratum? The etymological identification of items of Dravidian origin (**9.2**) is a necessary, but not a sufficient, criterion. It must also be shown that the spatial distribution of these elements is such that they cannot be accounted for merely by contact with adjacent Dravidian-speaking regions, or by historical contacts of the kind that occurred when, for example, the area known as Marathvada was part of the dominions of the Nizam of Hyderabad (**9.3**). In addition, it will be necessary to eliminate the possibility that the elements in question were *already current* in the speech of those Indo-Aryan speakers who are presumed to have entered Maharashtra (the Allchins' "immigrants from the north"). Only when these points have been demonstrated with a reasonable probability can we return to the discussion of the Allchins' proposed link between Dravidian speakers and the Malwa or pre-Malwa archaeological assemblages (**9.4**).

9.11. Maharashtra: political divisions

The political divisions of Maharashtra are shown in Figure 9.1. The state was created in 1965 on a linguistic basis, and thus is presumed to include all those areas in which the majority of the inhabitants claim Marathi as their mother tongue. Figure 9.2A shows, in schematic form, the former district organization on which the present analysis is based, giving an approximate idea of the relative positions of the districts.[2] The population figures are based on the Indian Census of 1991 (see India, Government of 1999).[3]

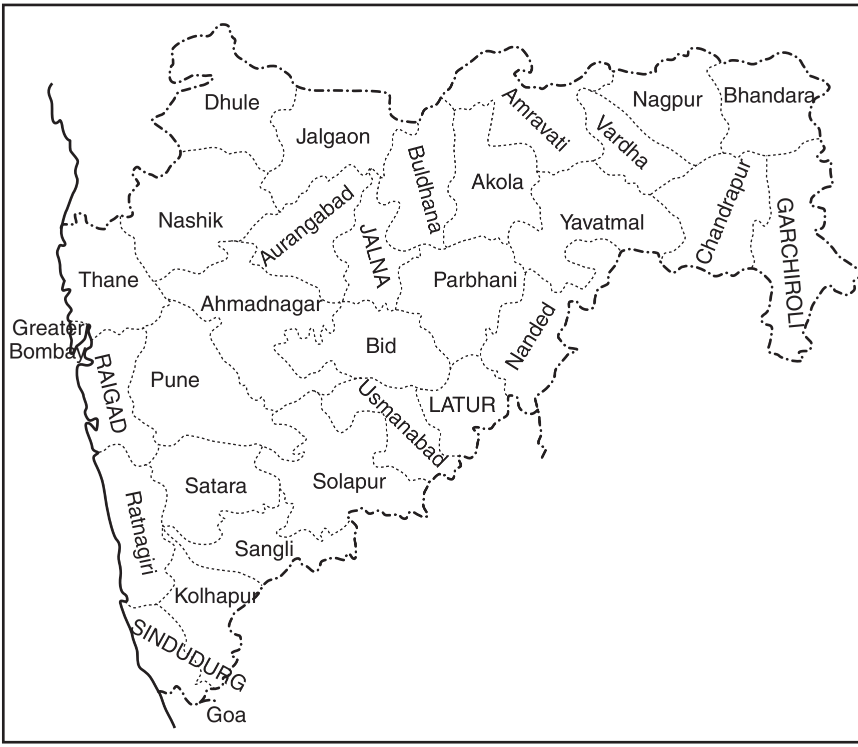


Figure 9.1 Districts of Maharashtra (new districts shown as follows: LATUR).

The state as a whole can be divided into five traditional regions, as indicated in Figure 9.2B:

(1) Varhad (Marathi **varhāḍ**, sanskritized as Vidarbha and anglicized as Berar);
(2) Khandesh (M. **khāndeš)**, the region of the Khandeshi or Ahirani language;[4]
(3) Konkan (M. **ko(ŋ)kaṇ)**, the coastal strip lying between the western ghats and the Arabian sea;
(4) Maharashtra (M. **mahārāṣṭra)**, also known as the **deš**, the plateau region traditionally most closely associated with Marathi culture (this region will be referred to in this chapter as Maharashtra-desh); and
(5) Marathvada (M. **marāṭhvāḍā**), the Marathi-speaking regions which were formerly part of the state of Hyderabad.

9.12. Geography and prehistory

Physically, Maharashtra consists of two major divisions: the plateau, which is part of the northern part of the Deccan tableland, and the coastal strip along the Arabian Sea,

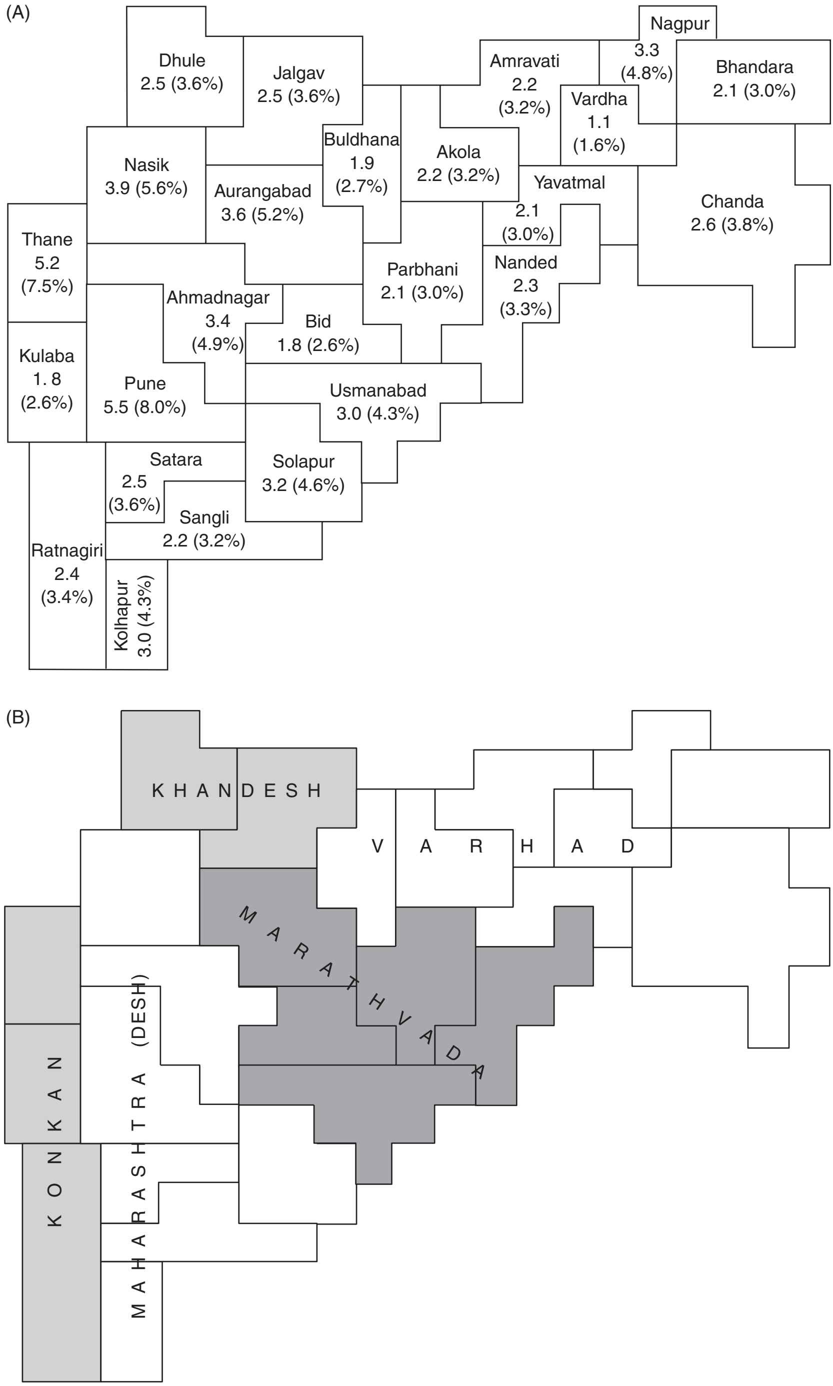


Figure 9.2 (A) Earlier districts of Maharashtra (schematic); District populations (in millions) and percentage of state population from Census of India 1991. (B) Regional divisions of Maharashtra.

known as the Konkan. These two segments are separated by the mountains of the Sahyadri range, which extends southward into the state of Karnataka and separates the drainage of the Godavari and Krishna Rivers, which drain toward the Bay of Bengal, from the rivers of the Konkan. To the north of the Sahyadris, the Narmada and Tapi River systems drain to the west. The eastern districts of Varhad are in the lower-lying central area of India beyond the Deccan plateau proper.

Settlements in the Deccan during the Palaeolithic period were confined to the river valleys, spreading to the upper levels of the plateau during the Mesolithic (Sankalia 1974). In a discussion of the Chalcolithic cultures of the Deccan, S. B. Deo points out that the Malwa Culture of the northern and central Deccan (dated in the range of 2300 (±70) BCE at Navdatoli on the Narmada, down to 1025 (±170) BCE at Inamgaon on the Ghod, a tributary of the Bhima)[5] show the presence of "pre-Chalcolithic Neolithic elements" which are presumably to be traced to Neolithic cultures of the Andhra-Karnatak region which flourished between 2295 (±155) and 1360 (±210) BCE. There are no purely Neolithic sites in the region of these Chalcolithic cultures (Deo 1982: 17–18). Deo also notes that the Malwa culture itself seems to have flourished first in central India (the present Madhya Pradesh) and "descended in the Deccan at a later period, i.e. in about the 16th century" [BCE] (1982: 23). The subsequent Jorwe Culture as a whole ranged between 1400 and 700 BCE (1982: 17).[6] The sites of both cultures are found primarily in the western part of the plateau in the river systems of the Godavari and the Bhima (a tributary of the Krishna). Some sites show evidence of contact with a Megalithic Culture, which flourished during the sixth and seventh centuries BCE, largely in the area of Vidarbha (**varhāḍ**) (1982: 35–44).

The Konkan region has produced practically no Palaeolithic or Mesolithic remains, other than somewhat isolated surface sites. Nor have any Chalcolithic or Megalithic remains been discovered in the region, even though Harappan and late Harappan settlements are found on the coast to the north of the river Tapi, while Bharuch on the Narmada shows possible traces of Chalcolithic artifacts, and the same site in period II (latter half of the first millenium BCE) represents the first Early Historical settlement on the coast. The Early Historical period is also represented at Sopara in present Thane Dt, where some fragments of Ashokan edicts have been found (Bloch 1950).

In the western Deccan, Early Historical settlements occur from about the fifth century BCE, for example at Nasik (Sankalia and Deo 1955). Many of these sites represent reoccupations of Chalcolithic sites in the Godavari and Bhima basins. A number of these sites were linked during this period to Konkan ports through various passes along the Sahyadris, and Buddhist caves are located at the heads of most of the important passes. During the period of the Satavahanas (OIA **sātavāhana-**), beginning around 100 BCE, the number of sites in the western Deccan increased, with a change in house construction techniques from the earlier mud huts to houses of brick with tiled roofs (Deo 1974), and a number of other differences which have led some scholars to speak of urbanization of sites during this period. Sites such as Daimabad and Inamgaon were large villages during the Chalcolithic period (Misra 2001: 516).

Historical sources do not enlighten us as to the linguistic or ethnic makeup of the Deccan before the medieval period. The area south of the Vindhyas was referred to in Sanskrit literary works as **dakṣiṇāpatha-** ('southern road'), which according to Bhandarkar referred to a region "almost identical with the country called Mahārāshṭra or the region in which the Marāṭhi language is spoken" (1927: 3–4). Bhandarkar suggests that, as the grammarian Panini (fifth century BCE) shows no acquaintance with India south of the Vindhyas but his commentator Katyayana (third century BCE) does so, the exploration of the southern region by northern speakers of Indo-Aryan must have begun sometime between the dates of these two grammarians, that is, in the late first millennium BCE (ibid.).

Under the name Aparanta (OIA **aparānta(ka)-** 'western extremity'), the northern Konkan is referred to in the Mahabharata, Kautalya's Arthashastra, and other literary works. Sopara (OIA **šūrpāraka-**) is mentioned both in Buddhist **jātaka** tales and in the Mahabharata. The grammarian Katyayana mentions the town of Nasik (OIA **nāsikya-**) on the Godavari, as well as Mahishmati (modern Maheshwar) on the Narmada. Paithan (M. **payṭhaṇ** ← OIA **pratiṣṭhāṇa**) on the Godavari is mentioned in Sanskrit works, as well as in the *Periplus of the Erythrean Sea*, which dates to the first century CE (Schoff 1974). Thus, historical sources suggest that speakers of Indo-Aryan began to appear in Maharashtra any time after the mid-first millennium BCE, whereas other evidence suggests that speakers of "outer" Indo-Aryan might have arrived up to a millennium earlier (see **6.4**). The existence of Sanskrit inscriptions from the late centuries BCE does not in itself guarantee anything about the identity of the creators of the inscriptions, since (e.g.) the Satavahanas, who are believed to belong to the "Andhra **jāti**", produced all their inscriptions in Sanskrit. Furthermore, it is not entirely clear whether the term "Andhra" in this context refers to language, regional origin, ethnic identity, or some combination of these three, since the term was not associated with the Telugu language until the eleventh century CE.

Though the whole of Maharashtra is Marathi-speaking today, it is not clear how far back that situation can be extrapolated. Deshpande, quoting S. B. Joshi, suggests that the area we now know as Maharashtra may have been predominantly Kannada-speaking until the twelftth century CE.[7] According to this argument, "the separate Marathi identity gradually evolved until in later periods the Maharashtrians connected themselves more with the Sanskritic north than with the Dravidian south" (Deshpande 1979b: 102; v. also 128). As noted elsewhere, there is internal linguistic evidence in Marathi for the continuation of Dravidian influence up to the twelfth century CE (see **9.1**, and Southworth 1976b). Again, it is not clear whether we should take the term "Kannada" to refer explicitly to the Kannada language, the region of Karnataka, or to a particular social group.

9.13. Structure of Marathi place names

Place names in Maharashtra, as in most parts of South Asia, show a dominant pattern which can be described as consisting of 'Name ± Suffix' (sometimes

referred to as ‘Specific ± Generic’). Examples:

> Name only: **keḷ**, **goṭe**, **māl**, **bori;**
> Name + Suffix: **keḷ-vihir**, **keḷ-vali**, **goṭe-gāv**, **goṭe-vāḍi**, **cinca-kheḍ**, **māl-pur**, **bori-vli**.

Names and suffixes are largely – but not completely – distinct classes of forms.[8] Suffixes include:

(a) forms denoting types of settlements, such as **gāv** ‘town/village’ (← OIA **grāma**), **ner** ‘town’ (← OIA **nagara**), **vali** ‘hamlet’ (← Dravidian, see **9.21**);
(b) those denoting the presence of water, such as **vihir, jhara, cuvā** (all meaning ‘well’), **kuṇḍ** ‘pond’;
(c) features of the physical environment, such as **ṭek, ḍongar** ‘hill’, **kār** ‘black rock’, **goṇḍi** ‘cattle track’, **poy** ‘gully’, etc.;
(d) a number of other identifiable categories, in addition to a large number of forms of uncertain meaning and unknown provenance.

In addition, there is a set of *secondary suffixes*, which generally are either related to administrative features of the settlement (e.g. **anāmat** ‘government land’, **khās** ‘private’), or else serve to distinguish two otherwise identical names (e.g. **khurd** ‘smaller/ younger’ – **budruk** ‘larger/ older’, **tarph(e)** X ‘near or toward X’, **urph** X ‘also known as X’, **pratham** X ‘formerly X’; examples: **cinca-vali tarph atoṇe, rāvaṇ-vāḍi urph tembhḷi**. Other minor patterns occasionally occur.

Suffixes in Marathi place names include, in addition to *tadbhava* (inherited) forms such as **ner** and **gāv** and *tatsama* forms (learned borrowings from Sanskrit) such as **nagar** and **grām**, forms of foreign origin such as **kasbā, ābād, haveli** of Perso-Arabic derivation, and **roḍ** (road), **jankšan** (junction), **ṭāun** (town) of English origin. Names form a much larger class, in which many members appear to be proper names without any other discoverable referential meaning, though this class also includes many common nouns, such as names of trees. (See additional discussion in **9.3**.)

The source of data on Maharashtrian village names used for this chapter is N. G. Apte’s *Mahārāṣṭra Rājya Grāmasūci* [“Maharashtra state village-list”] (Apte 1967), hereinafter referred to as the *Gramasuci*.

9.2. Suffixes of probable Dravidian origin

Toponymic studies in India have focused primarily on the *suffixes* of place names, and thus it is possible to draw on a sizeable body of work dealing with these suffixes. The *Gramasuci* contains upwards of 300 suffixes, the majority of which are of uncertain origin. The following discussion focuses on eight suffixes of probable Dravidian origin which are found in ancient as well as modern place names in Maharashtra.

9.21. vali

The suffix **-vali/-vli**[9] (as seen e.g. in **nā-vali, bori-vli**), has long been recognized as probably derived from Dravidian (see **pallī** in **3.22A(4)**, also CDIAL 7972). Its presumed Dravidian source is PD ***paḷḷi** 'hamlet, village', which occurs in all the literary Dravidian languages, as well as Parji, as a place-name suffix (DEDR 4018). This suffix is in turn derivable from PD1 ***paḷḷi** 'group, aggregation', possibly originally 'family group' (cf. Kudux **erpa palli** 'household' [**erpa** 'house']). It is not improbable that this suffix contains the same PD root **paṭ-** 'fall, lie' found in **paṭṭi**, etc. (**9.23** below). This word is not found in early Sanskrit literature, but occurs (in the form **pallī-** or **palli-**) in the Kathasaritsagara; according to D. D. Kosambi, in this work it designates only the "hut-clusters of Bhils" (1956: 247). In the Jaina Prakrit Uttaradhyanasutra, whose oldest parts may go back as far as 300 BCE, the word **corapallī-** means 'den of thieves'. Sankalia points out that though "the underlying idea in [Sanskrit] literature is thus of insignificance and unrespectability", in the inscriptions it appears to be a general word for a small settlement; furthermore, a **pallī** which grows beyond a certain point can become a **grāma** (see below), with the latter suffix either following or replacing the former (Sankalia 1949: 53). In Maharashtra forms of this suffix begin to appear in inscriptions from the period of the Calukyas of Badami (fifth to eighth centuries CE). In a few cases it is possible to establish a link between inscriptional names ending in **-palli** and modern place names in **-vali** or its equivalent: for example, Cikhyaliapallikā (Mulay 1972: 138) probably = Cikhaḷoli (Thane Dt).

There is a possibility of confusion between this suffix and another, which appears in the earliest inscriptions as **valli** or **vaḷḷi** (Mulay 1972: 139). This also appears to be a suffix of Dravidian origin, and is known in Dravidian place names in Tamilnadu, Kerala, and Karnataka (George 1986: 258–9). The Dravidian source is uncertain, since the only similar item in DEDR is 5316 **vaḷḷi** 'creeper'. In Maharashtra, the two suffixes show similar geographical distribution (see below). It is not possible to tell in all individual cases which is the original suffix involved, since the older **-p-** changed to **-v-** in Marathi and other Indo-Aryan languages before the beginning of the Old Marathi period (i.e. before the eleventh century CE). Some of the cases listed by Mulay as **valli/vaḷḷi** (1972: 139–40) probably represent original **paḷḷi**, since their modern forms in Karnataka show **h**, the usual modern Kannada reflex of older **p**, whereas older **v** normally changes to **b** in Kannada. In any case, we can probably assume Dravidian origin for both of these forms.

As an isolated word, the expected reflex of a Dravidian **paḷḷi** in Marathi would be **pāl(i)**. As an unstressed suffix we would expect **vali, ul(i)**, or **ol(i)**. One might also expect that restressed varieties of the suffix might occur (along the lines of BIRmingHAM or NEW-TOWN ← Newton), namely, **vāl(i), pāl(i)**.[10]

9.22. Geographical distribution of -vali *and variants*

Figure 9.3 shows the distribution of place names in **-vali** in Maharashtra. (This figure is based on the total list of villages in the *Gramasuci*.) Here the regions of

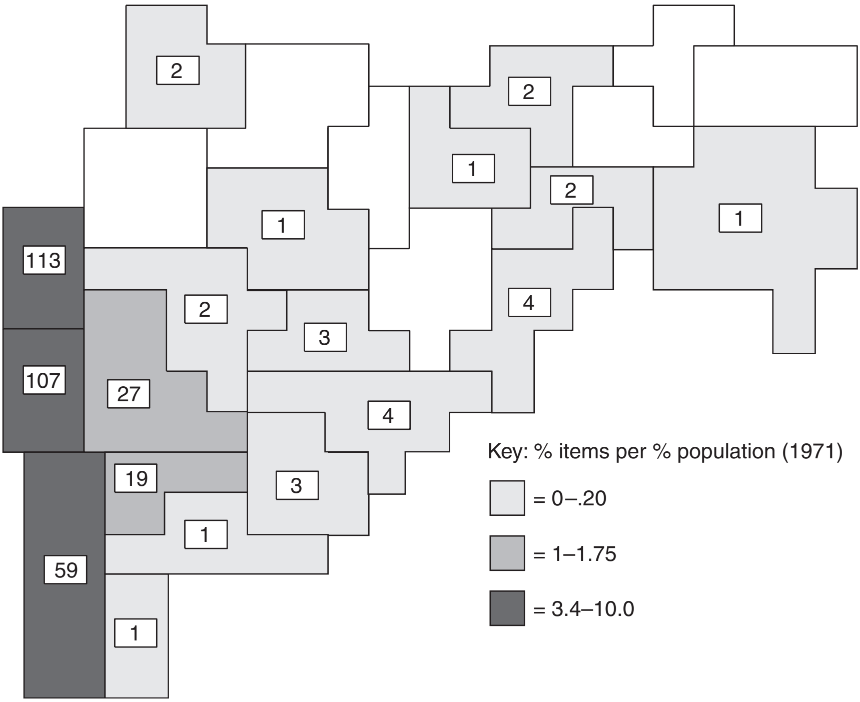


Figure 9.3 Distribution of village names in -**v(a)li**.

Source: Apte 1967 (100 percent sample).

Konkan and (part of) Maharashtra-desh show the highest incidences of this particular suffix.

A sharp contrast to Figure 9.3 is shown by Figure 9.4, which shows the distribution of names containing the suffix **gāv/gā̃v** (← OIA **grāma**, an Indo-Aryan suffix of great antiquity, derived from Proto-Indo-European). Thus the distribution of **gāv** is presumably linked to settlements of Indo-Aryan speakers. Though suffixes derived from OIA **grāma** occur in the areas where Kannada is now spoken, and which as far as we know were never Indo-Aryan-speaking, they are much rarer than in the Marathi-speaking areas, and the modern forms are usually different, namely **gām, gāmi**, as opposed to Marathi **gāv**. Whereas Figure 9.3 shows the greatest frequency of the suffix **-vali** in Konkan and parts of Maharashtra-desh, Figure 9.4 shows **gāv** to increase in frequency toward the northeast and decrease toward the southwest.

Figure 9.5 shows the occurrences of the suffix **-grāma** in inscriptions, based on Mulay (1972). Apart from the near-absence of cases in eastern Maharashtra, that is, Varhad, the proportions of cases in the other districts do not differ greatly

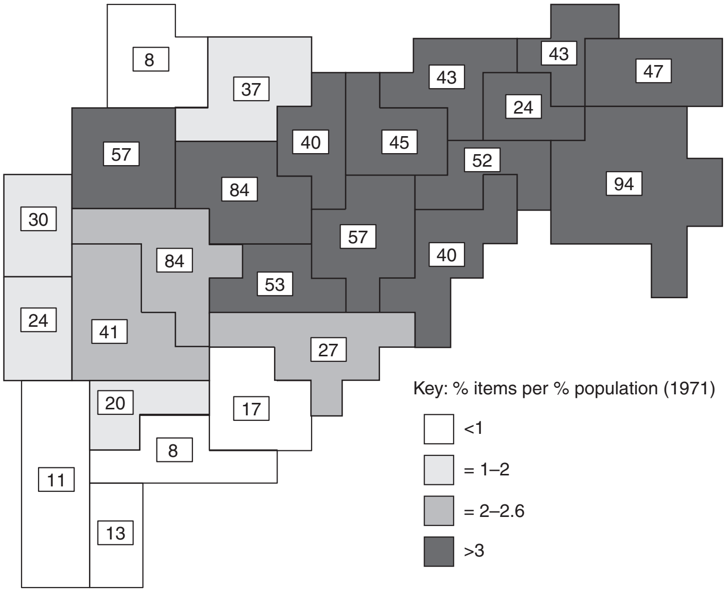


Figure 9.4 Distribution of village names in **-gāv**.

Source: Apte 1967 (20 percent sample).

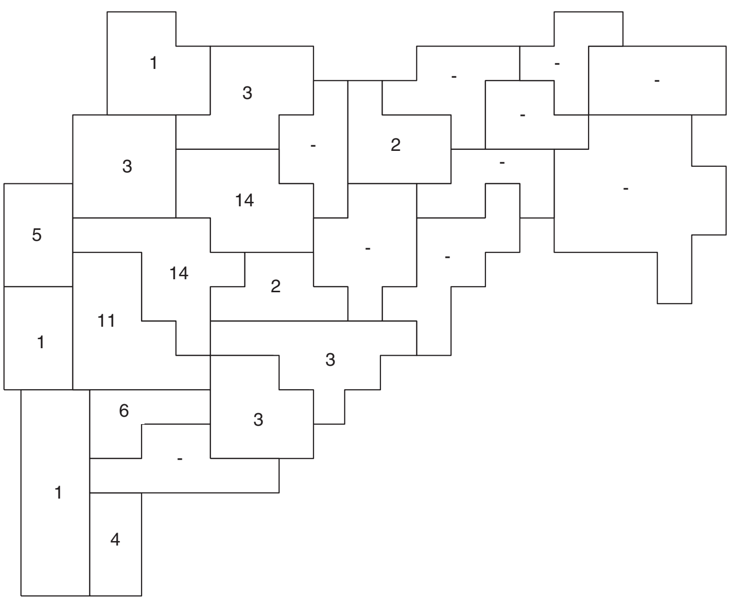


Figure 9.5 Inscriptional occurrences of village names in **-grāma**.

Source: Mulay 1972: 124–8.

from those shown in Figure 9.4. It is not clear whether the absence of cases in the east represents the true situation prior to the twelfth century CE, or if it is merely the result of the general lack of inscriptional place names found in the Varhad area at that period. (See further discussion in **9.4**.)

Since **-vali** is considered by most authorities to be of Dravidian origin, its high frequency in this region (coupled with the low frequency of the suffix **gāv**) strongly suggests the likelihood of the earlier presence of Dravidian speakers in Konkan and southern Maharashtra-desh. The inscriptional evidence, meager though it is, supports this conclusion: the two cases of names in **palli** cited by Mulay (1972: 138) are both in Thane Dt of Konkan. We may add to this the fact that names in **-vali** also occur in the inscriptions of Gujarat (Sankalia 1949: 53–4), which is also part of Trautmann's area of "Dravidian kinship" (Trautmann 1981; see **9.1**). Thus the region in which this suffix is found is continuous from Gujarat, through Maharashtra (Konkan and parts of Maharashtra-desh), into south India. (See **9.4–9.6** for further discussion.)

Suffixes in **-oli**, pobably also ←Dravidian **paḷḷi** or **valli**, show a somewhat different distribution, as indicated in Figure 9.6. Here the two southern districts

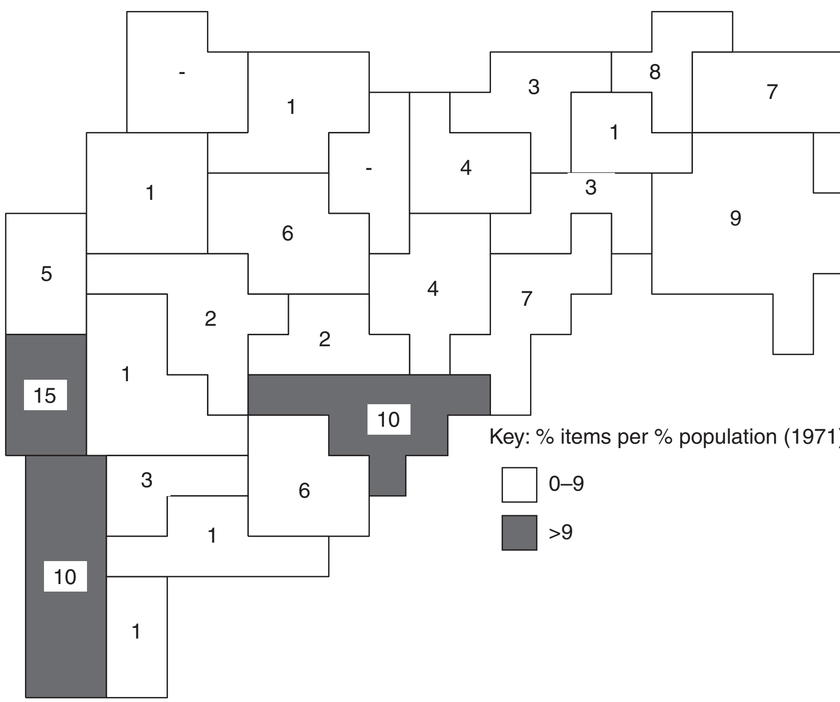


Figure 9.6 Distribution of village names in **-oli**.

Source: Apte 1967 (20 percent sample).

of Konkan (Kulaba and Ratnagiri) still show the highest incidence, along with Usmanabad Dt of the Marathvada region, but several other districts are not far behind – and statistically not significantly so. There is a possibility, of course, that some cases of **-oli** derive from older **valli** rather than **palli**, as noted earlier; the older distribution of **valli**, as far as we can tell from the few cases found in inscriptions, seems to have been not very different from that of **palli**: see Figure 9.7.

If we combine all cases of **vali, oli**, and other possible reflexes of **valli** or **paḷḷi** (**ali, uli, li**) from the 20 percent sample of the *Gramasuci*, we get the picture shown in Figure 9.8, which still shows the preponderance of cases in Konkan, though Usmanabad Dt of Marathvada and Chanda Dt of Varhad are not far behind.

It is not immediately apparent why the distribution of **-oli** is different from that of **-vali**. One possibility would be a regional difference in sound changes, such that the change **av(a)** → **o** before consonant is less frequent in Konkan than elsewhere. No such difference is mentioned in the literature, but some data from the

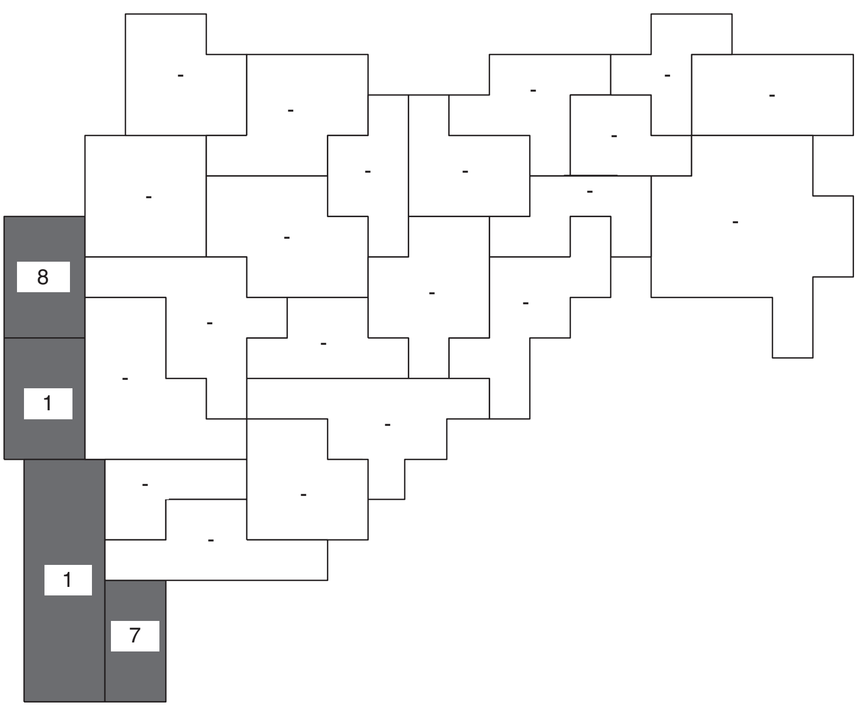


Figure 9.7 Inscriptional occurrences of village names in -vallika, -vulika, -al(l)i, -uli.
Source: Mulay 1972: 139–40, 154–5.

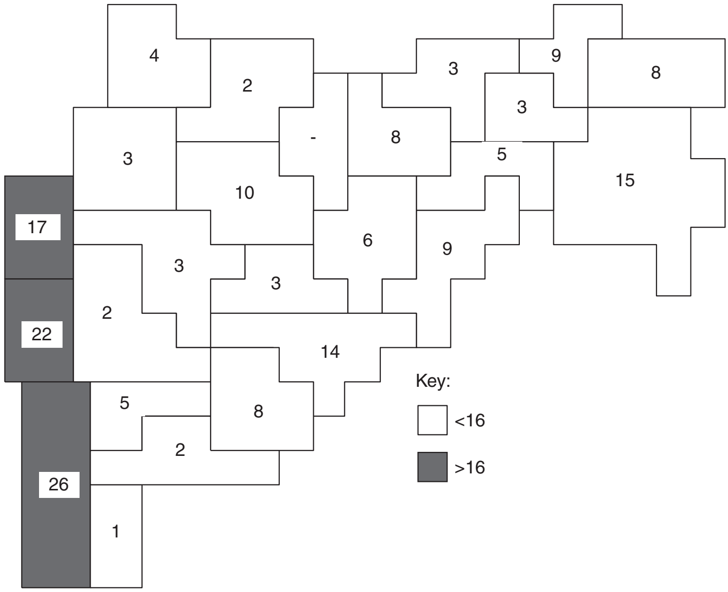


Figure 9.8 Distribution of reflexes of -valli and -paḻḻi.

Source: Apte 1967 (20 percent sample).

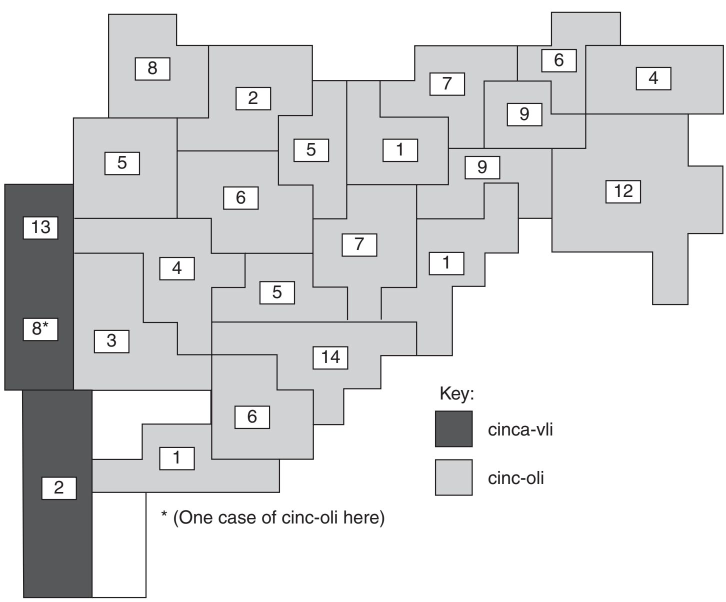


Figure 9.9 Distribution of village names **cinca-vli** and **cinc-oli**.

Source: Apte 1967 (100 percent sample).

present study suggest that there may be such a difference. Figure 9.9 shows the distribution of the village names **cinc-avli** and **cincoli**, indicating that the former occurs only in Konkan, while the latter occurs widely in the rest of the state. On the other hand, there are eight occurrences of **pimpaḷ-oli** (**pimpaḷ** = 'pipal tree'), of which seven are in Konkan and one in Maharashtra-desh, and no cases of **pimpaḷ-avli**.

It is possible that all these developments can be accounted for in terms of regionally conditioned sound change (**1.21C**). On the other hand, it may be that in areas where reflexes of Dravidian **paḷḷi** survived as a non-suffixal morpheme, either as a noun meaning 'village, hamlet' or as the first part of a village name, the presence of the full variant **pāl(i)** (see above) may have served to make the reduction of **-v(a)li** to **oli/uli/(a)li** less frequent, since speakers' internal grammars would have included a morphophonemic rule relating **pāli (←pallī)** to **v(a)li**, whereas this would not have been the case in areas where **pāli** did not occur independently. Figure 9.10 indicates that the area in which **-oli** is less frequent, that is, Konkan, is the area which shows the highest

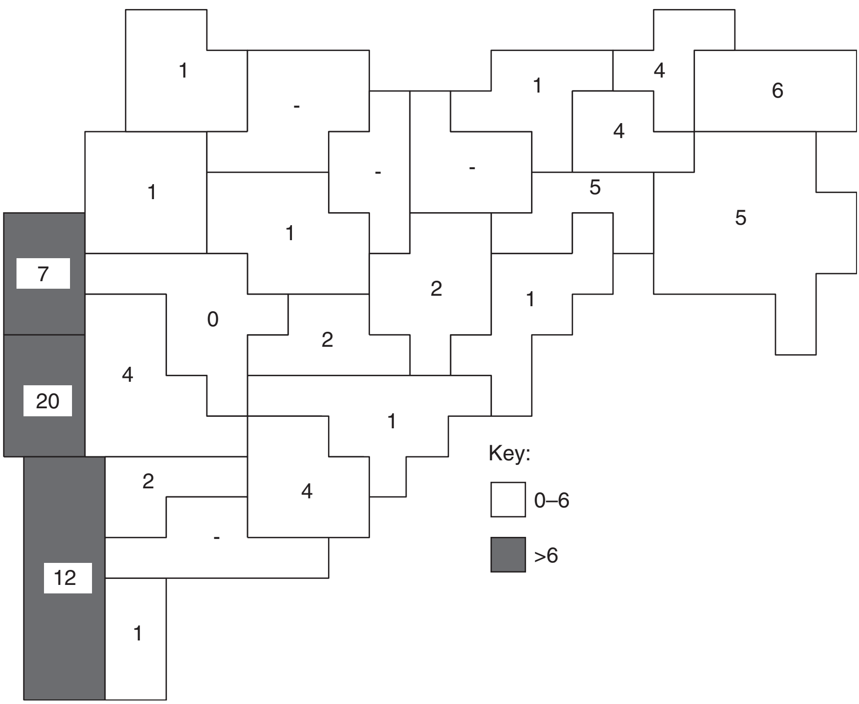


Figure 9.10 Distribution of **pāl(i)** as village name.

Source: Apte 1967 (100 percent sample).

incidence of village names in **pāli** (±suffix), as well as other names beginning in **pāl-**.[11]

9.23. vāḍā/vāḍi

S. B. Joshi (1951) has discussed a number of suffixes which he derives from a Dravidian **paṭṭi**. The Dravidian items in question are the following:

(1) (a) PSD ***paṭṭi/paṭṭu/paṭṭa** 'cowstall; settlement';
(b) Ta. **paṭṭiṇam** 'town, seaport' (DEDR 3868).
(2) PSD ***pāṭi/pāṭu/pāṭa(ka)m** 'hamlet, part of village' (DEDR 4064).

The semantics of these two items in Dravidian cover the entire range from places for bedding down animals, through pastoral camps and small agricultural settlements, to hamlets incorporated into villages, to towns and larger settlements.[12] Thus as Joshi suggests, these words appear to reflect the entire history of human habitation beginning with the temporary camps of nomadic or semi-nomadic cattle keepers, and including the transition from cattle shelters to human settlements, as well as the later incorporation of pastoral or agricultural hamlets into larger village complexes. (Ta. **pāṭam** means, among other things, 'street of herdsmen'.) It seems reasonable to derive these forms, as Joshi suggests, from a PD ***paṭ-** 'fall, lie, sit (of humans or animals)' (DEDR 3852), in accordance with the probable original meaning of these suffixes, that is, 'lying place (for humans and/or animals)'. Note that these forms (i.e. **paṭṭi**, **pāṭi**, and **paṭṭaṇam**) all occur as place-name suffixes in the four literary Dravidian languages investigated by K. M. George (1986: 197–8, 206–8, 195–6).

The OIA word **paṭṭaṇa** occurs in Kautilya and in the MBh in the meaning 'town'. Mulay's investigation of place-name suffixes in the Deccan shows a number of cases of the suffix **paṭṭi** in Karnataka, but none in Maharashtra (Mulay 1972: 143). Note that there is a change of **p** → **h** in Kannada in the tenth century CH, so that the modern forms appear in Kannada-speaking areas as **haṭṭi** (Mulay 1972: 109).

As isolated words, the three Dravidian forms given before would be expected to appear in Marathi as (1a) **pāṭ(i)**, (1b) **pāṭaṇ**, (2) **pāḍ(i)**. As unstressed suffixes, we would expect (1a) **vaṭ(i)/oṭ(i)/uṭ(i)**, (1b) **vaṭaṇ/oṭaṇ/uṭaṇ**, (2) **vāḍ(i)/oḍ(i)/uḍ(i)**. Restressed varieties of the suffixal forms would yield (1a) **vāṭ(i)**, (1b) **vāṭaṇ**, (2) **vāḍ(i)**.[13] While many or most of these forms can be found as place-name suffixes in Maharashtra, the only ones which occur with any frequency are **vāḍi** and a form with **-ā**, **vāḍā**, both of which also occur as independent words in Marathi, and are derived by Turner from an OIA **vāṭa** 'enclosure, fence' (MBh), **vāṭī** 'enclosed land' (BhP), **vāṭikā** 'enclosure, garden' (Kathas.), which he believes to be derived in turn from an early MIA form ***var-tra** 'enclosure', formed on the root **vṛ** 'cover'(CDIAL s.v. **vāṭa**, an etymon which includes a large number of NIA words such as B. **bāri** 'garden').[14]

However, on semantic grounds it is not clear that Marathi **vāḍā** and **vāḍi** are derived (solely) from Sanskrit **vāṭa-** etc. Molesworth (1975) defines these words as follows:

vāḍā: (1) large edifice, (2) quarter of a town, (3) division of a village, (4) enclosed space, compound, (5) pen or fold;
vāḍi: (1) enclosed meadow or garden, paddock, (2) cluster of agriculturalists' huts, hamlet, (3) division of the suburban portion of a city.

Thus, both of these Marathi words appear to have part of their meanings derivable from Sanskrit **vāṭa-/vāṭī-** 'enclosure', and part derivable from Dravidian **paṭṭi** 'settlement' or **pāṭi** 'hamlet, section of village'. Note that Marathi also has a word **pāḍā**, whose meanings include 'a hamlet or cluster of houses of agriculturists' and 'a ward or quarter of a town'.[15] Thus it seems likely that the meanings 'hamlet' and 'division of a village or town' derive from the Dravidian **paṭṭi** and **pāṭi**, while the meanings 'enclosure', 'garden', 'compound' etc. derive from MIA ***var-tra** or the like. If this is the case, then the occurrence of

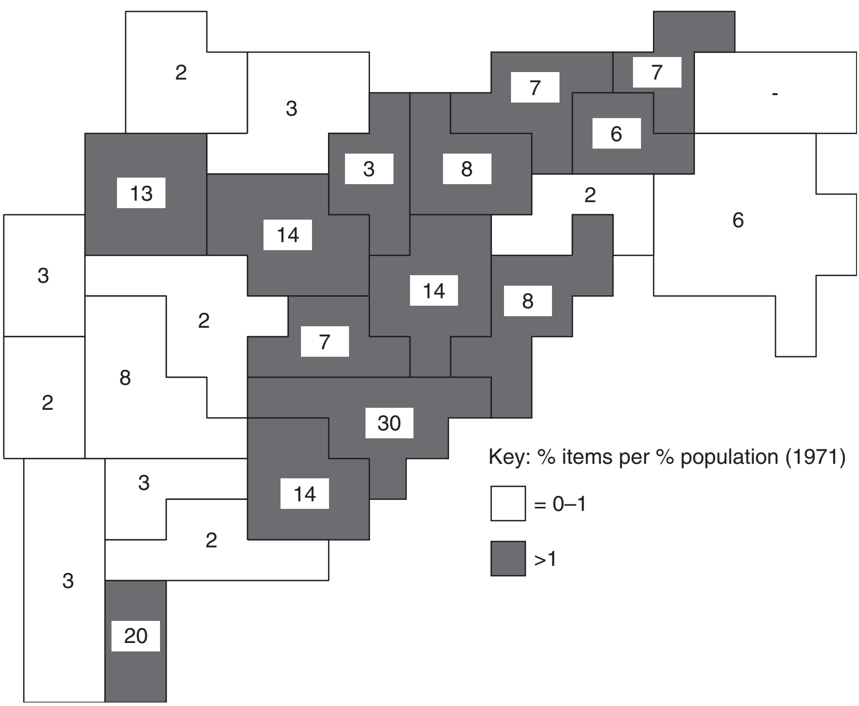


Figure 9.11 Distribution of village names in **-vāḍi**.

Source: Apte 1967 (20 percent sample).

vāḍā, etc. as place-name suffixes in Maharashtra would seem to be traceable to Dravidian.

The geographical distribution of these suffixes may throw further light on this question, though it fails to resolve the question of origin definitively. The most frequently occurring suffix in this group is **-vāḍi**, whose distribution is shown in Figure 9.11. This pattern appears to be closer to that of Figure 9.4 (**gāv**) than to that of Figure 9.3 (**vaḷi**). If the assumption is correct that the distribution shown in Figure 9.4 relates to areas of primarily Indo-Aryan-speaking settlements, it would imply that **vāḍi** may also be primarily Indo-Aryan in origin, that is, it may derive from OIA **vāṭ-** etc. rather than directly from any of the Dravidian forms cited earlier.

Figure 9.12 shows the distribution of **vāḍi** as a specific, that is, as the first part of a village name. This distribution does not really match either that of 9.4 or 9.3, but perhaps the number of cases is too few to be significant. Figure 9.13, on the other hand, which depicts the distribution of **pāṭ** (±suffix) as a village name, shows a distribution similar to that of Figure 9.3, suggesting that these cases may well derive from Dravidian **paṭṭi** (since cases beginning with OIA **v-** do not enter the picture here).

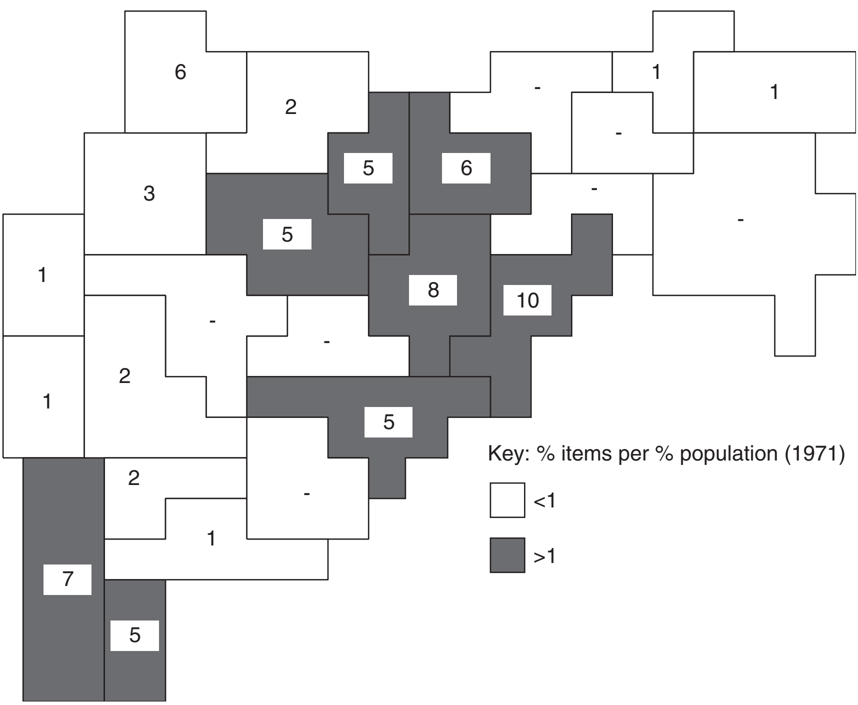


Figure 9.12 Distribution of **vāḍi** (±suffix) as village name.

Source: Apte 1967 (100 percent sample).

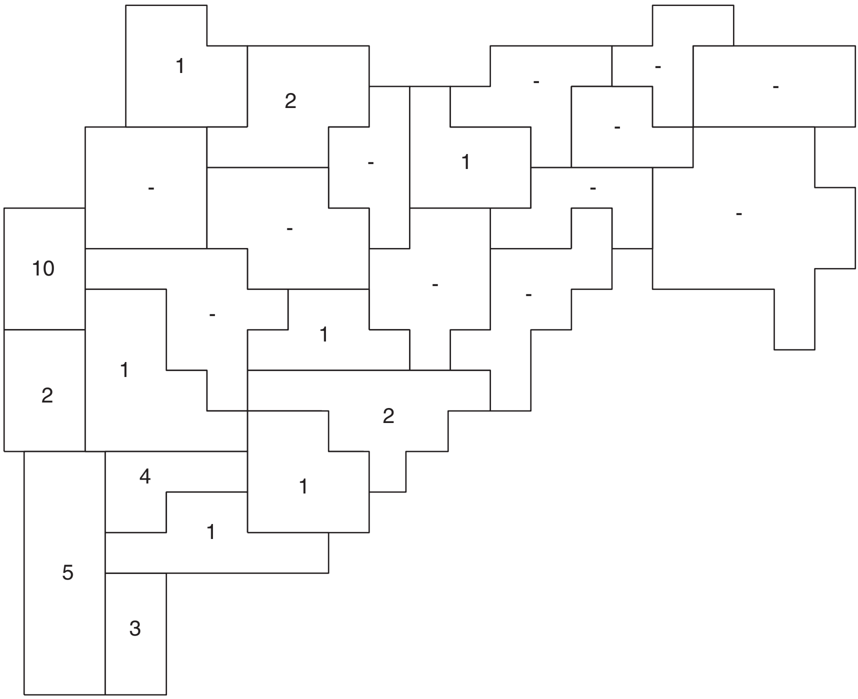


Figure 9.13 Distribution of **pāṭ** (±suffix) as village name.

Source: Apte 1967 (100 percent sample).

The results of this inquiry into the origins of the Marathi suffixes **-vāḍā/vāḍī** are somewhat ambiguous, but perhaps sufficient to show that the history of these suffixes is not the same as either that of **-vali** or **-gāv**. Taking all of the above evidence into account, it seems most probable that Indo-Aryan and Dravidian (and possibly other)[16] elements are combined in the history of this group of suffixes, though it is impossible at present to be more specific.

9.24. Other suffixes

The suffixes discussed next are of probable Dravidian origin, and though less frequent in occurrence than those discussed before, may throw some additional light on the question.[17]

9.24A. ur

The PD ***ūr** (DEDR 752) 'settlement, habitation' (orig. 'house'?) is one of the oldest words in Dravidian for a settlement (see also **8.20**). It occurs as

a place-name suffix in all four of the literary Dravidian languages (George 1986: 96–8). Mulay notes that it is found frequently in inscriptions in Karnataka, but cites only two cases from Maharashtra, both in Konkan: **Nerur** in Ratnagiri Dt and **Naur** (modern **Nahur**) in Thane Dt (Mulay 1972: 135–6). A problem with this suffix is that its modern (and to some extent, even its earlier) occurrences are apt to be confounded with the reflexes of OIA **pura-**, a suffix of great antiquity in Indo-Aryan; because of the change **-p-** → **-v-**, both suffixes would end up as **-(v)ur** in Marathi, as Mulay has noted (1972: 134). Figure 9.14 shows the distribution of village names in **-ur** based on a 10 percent sample from the *Gramasuci*, suggesting that there is no significant distributional pattern.

9.24B. kal

This suffix is found in place names in the four southern states (George 1986: 114–16). Mulay finds 11 cases of inscriptional names ending in **kal**, all in Karnataka (1972: 150). The suffix is found in a few cases in modern place names

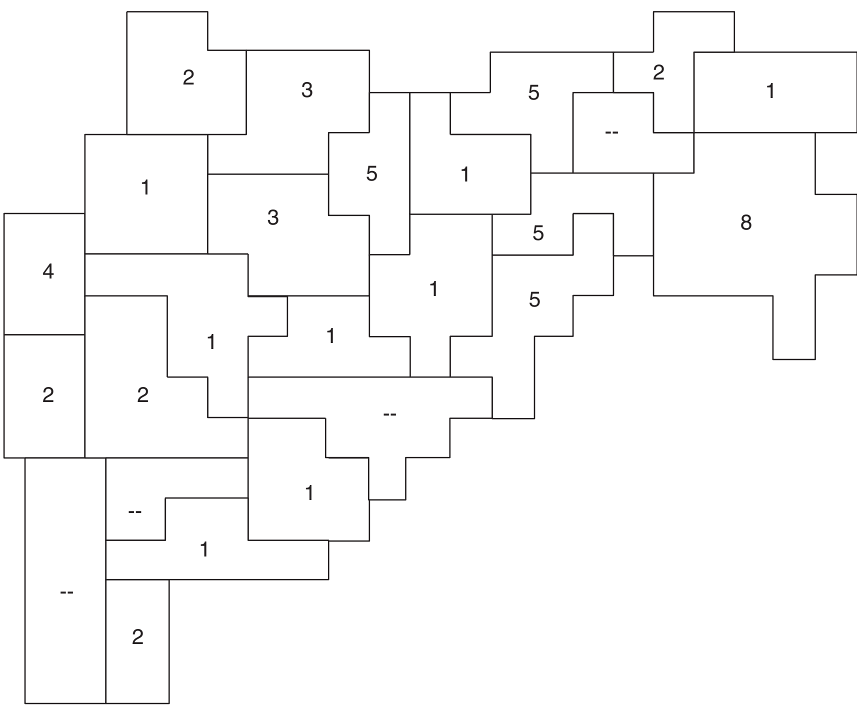


Figure 9.14 Distribution of village names in **-ur**.

Source: Apte 1967 (10 percent sample).

in Maharashtra (1972: 170). Derivation from PD ***kal** 'stone' (DEDR 1298) is a possibility.

9.24C. kuṇḍ

This suffix is found in inscriptions in Karnataka (11 cases) and Maharashtra (1 case), and in modern place names in Maharashtra (Mulay 1972: 116, 148). It is found in place names throughout South India, and is probably to be derived from PD ***kuṇṭ** 'pond; hollow' (DEDR 1669). A possibly related item is the Marathi word **nigaḍe**, which occurs both as a place name and in the meaning 'the shrub *Vitex negundo*, found in the plains and lower hills of India, common near streams' (CDIAL 7308). A very plausible source for both the OIA and Marathi words would be a Dravidian ***nīr-kuṇṭ** 'water hole' (= PD ***nīr** 'water' [DEDR 3690a] + PD ***kuṇṭ**); see **3.22A(4)**.

Figure 9.15 shows **kuṇḍ** to be of very low frequency, both as a suffix and as a village name. To the extent that any pattern is discernible, Figure 9.15 appears

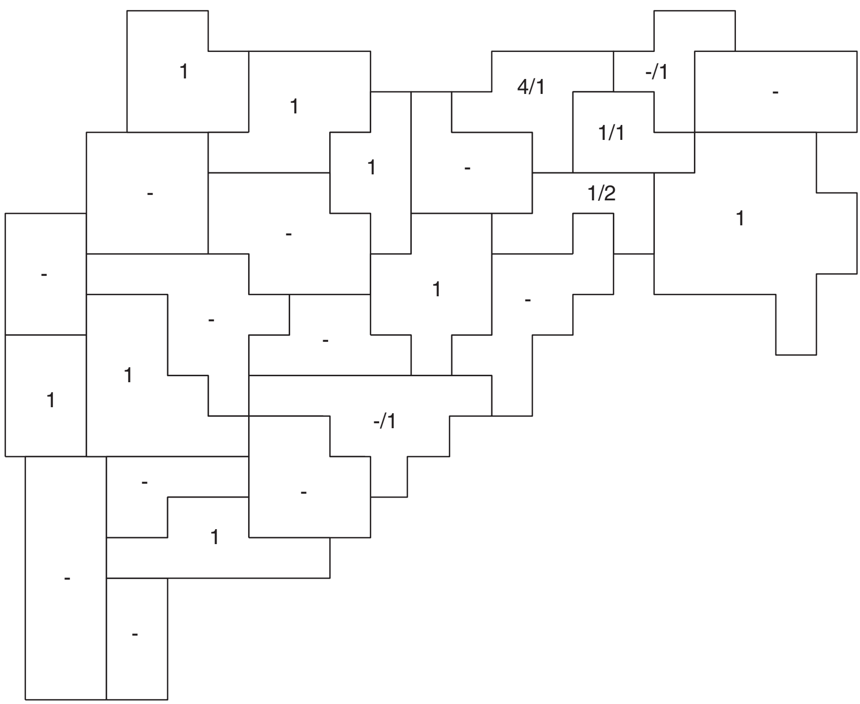


Figure 9.15 Distribution of **kuṇḍ** as suffix/as name.

Source: Apte 1967 (20 percent sample).

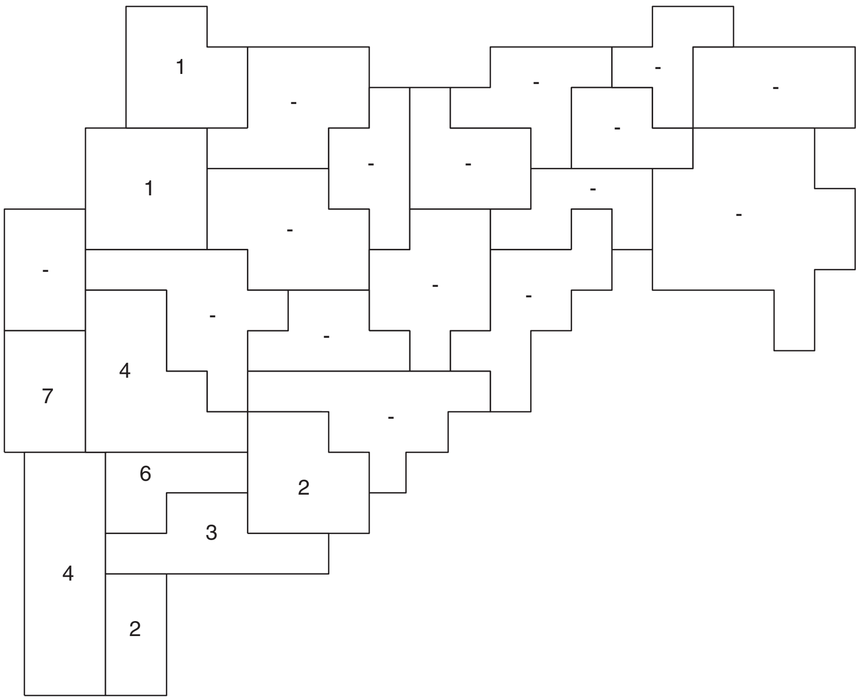


Figure 9.16 Distribution of **nigaḍe** and variants as village names.
Source: Apte 1967(20 percent sample).

more like Figure 9.4 (**gāv**) than Figure 9.3 (**vali**), which may reflect the fact that **kuṇḍ**, though probably of Dravidian origin, was borrowed early in OIA (see **3.22A4**). In contrast, the distribution of **nigaḍe** (and variants **nigaḍi, niguḍ(e)**), as shown in Figure 9.16, is similar to that of **vali**, suggesting that this word is probably of Dravidian origin and was not adopted by OIA speakers except as a suffixal element.

9.24D. kuppa

The suffix **kuppa(m)** is found in place names throughout South India, and is presumably to be derived from either PD ***kupp-** 'assemblage, heap' (DEDR 1731) or PD ***kupp-** 'hut' (DEDR 1732). Mulay notes a few names containing this suffix: for example, Belakuppe, Siriguppi, Juguḷakoppa, both in northern Karnataka and southern Maharashtra (1972: 121, 151).

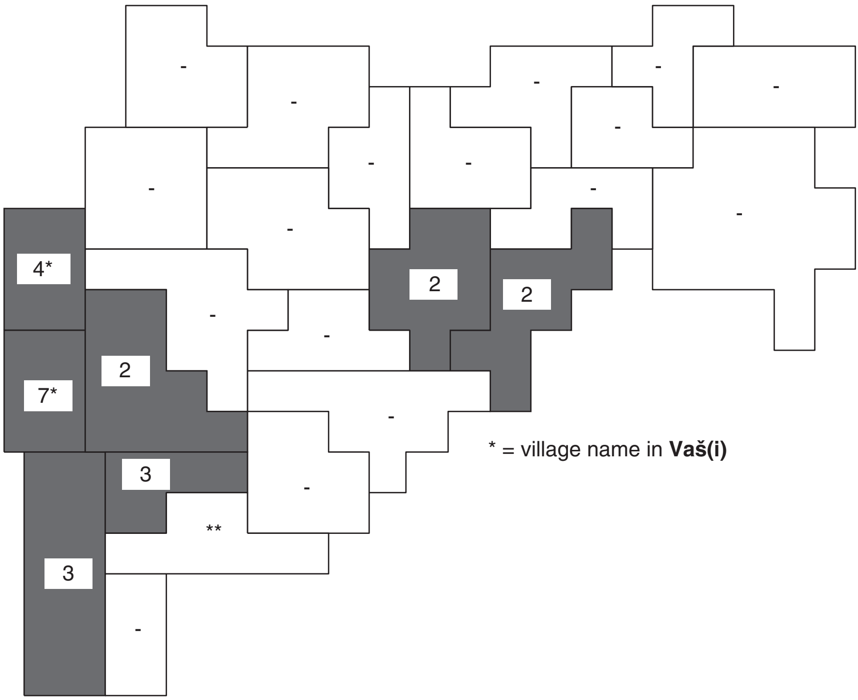


Figure 9.17 Distribution of village names in **-oši**, **-uši**, **-oše**, **-os**.

Source: Apte 1967 (20 percent sample).

9.24E. oši

Mulay (1972: 175) notes that this suffix is found mainly in the Konkan (92 of 95 cases), and suggests that it may refer to some "coastal feature". Derivation from SD ***vaci** '(flowing) water' (DEDR 5214) is a possibility. This suffix and variants (**-uši, -oše, -os(e)**) occur in 23 cases in the *Gramasuci*, distributed as in Figure 9.17; in addition, village names beginning in **vaš(i) …** (**Vaši**, **Vašoṇi**, **Vašāḷe**) occur in three districts, as indicated in the figure.[18]

9.24F. ge/gi

Mulay (1972: 161, 169) notes that this suffix is frequent in Karnataka and southern Maharashtra-desh (10 cases in Solapur Dt, 30 in Satara Dt), but extremely rare elsewhere. Possible source: PD ***keyi** 'field' (DEDR 1958, 1993). K. M. George (1986: 143) lists **kai** as a suffix of place names in Karnataka and Kerala.

9.3. Names occurring with the suffix -vali

If the suffixes discussed previously are indeed derived from Dravidian sources, and if the villages with names ending in these suffixes were in fact settled – or named – by an earlier population of Dravidian speakers, we would necessarily expect to find that at least *some* of the linguistic elements preceding the suffixes are derived, or derivable, from Dravidian origins. This section presents a detailed analysis of the names which occur with the suffix **-vali**. The other potentially Dravidian suffixes could, and should ultimately, also be examined from this point of view, but it will perhaps suffice here to examine this one suffix, for which the most detailed information is available.

The initial elements ("specifics") in Marathi place names fall into a small number of definable categories, principally the following:

(1) trees and plants (e.g. **umbar** 'udumbara tree', **cinc** 'tamarind', **nimb** 'neem', **pimpaḷ** 'pipal');
(2) physical features of the environment (**rāṇ** 'forest', **ḍongar** 'hill');
(3) names of castes or professional groups (**kumbhār** 'potter', **mahār** 'Mahar caste', **bāmaṇ** 'Brahman');
(4) other proper names (**ānand**, **gopāl**, **hanamant**, **khān**);
(5) animals (**gāy** 'cow', **gāḍhav** 'ass', **vāgh** 'tiger');
(6) village products (e.g. **dahi** 'curd', **sākhar** 'sugar' – these are rare, and therefore somewhat doubtful).

The names occurring with the suffix **-vali** include items from all these categories. The large majority of names are however not clearly identifiable semantically or etymologically, as is in fact the case with all the names in the *Gramasuci*. Of the approximately 139 names occurring with **-vali**, about 92 are of uncertain derivation. Of the remainder, 22 are more or less identifiable as Indo-Aryan,[19] and about 15 can be classed as probably Dravidian.[20] Another ten are *possibly* Dravidian, but the identification is less certain, for reasons explained below.

The identification of the first part of a name as (probably) Dravidian involves the following criteria:

(1) it must have a Dravidian etymology which is not only plausible, but more acceptable than any Indo-Aryan etymology;
(2) the meaning of the item must fit into one of the categories mentioned above;
(3) the distribution of the item must be predominantly in the area identified (above) as most likely to have been populated by Dravidian speakers in earlier times, that is, Konkan and the southern/western part of Maharashtra-desh.

Following are the 15 items which fit these criteria:[21]

Marathi Form		Dravidian (DEDR No., form, mg)	Distribution					
			KOK	*S. MAH.*	*N. MAH*	*KHA*	*MAR*	*VAR*
1	**āmbi**	177 **ampi** 'boat'[22]	22	4	5	—	1	2
2	**kaḍ**	1123 **kaṭa** 'male animal'	5	6	1	2	5	—
3	**kāši**	1432 **kāc-u/i** '*Acacia catechu*'	11	2	—	—	1	4
4	**kir**	1594 **kiṟu** 'small'	12	7	1	1	5	6
5	**kuḍ**	1655 **kuṭi** 'abode'[23]	28	11	—	1	2	10
6	**koṇḍ**	1864 **kuṉṟam** 'mountain'	44	10	3	—	8	12
7	**koḷ**	1867 **kōḷi** 'banyan/fig'	18	14	7	3	17	10
8	**koṇḍ**	2077 Kol. **gonḍa** 'Gond'[24]	35	4	7	—	11	10
9	**toṇḍa**	2928 **tōṇṭu** 'garden'	7	6	1	1	6	3
10	**nā**	3650 **nāy** 'dog'	9	4	—	—	—	—
11	**pāli**	4018 **paḷḷi** 'hamlet'	(see Figure 9.10 for distribution)					
12	**maj(i)**	4637 **manci** 'hemp'	5	8	—	—	—	—
13	**māni**	4780 **māṉ** 'deer'	8	—	—	—	—	—
14	**sir**	2628 **cīre** 'chironji tree'	46	32	7	12	4	7
15	**sāḍ**	2308 **caṭṭu** 'rock; flat'	6	6	2	3	—	—

One additional item encountered in the investigation of these names is **umbar**, which as a noun means 'the tree *Ficus glomerata*, glomerous fig tree'. This word, which is of Dravidian origin (see **3.22A3**) occurs as FN with the suffix **-vali/oli** in the following distribution:

KOK	*S. MAH*	*N. MAH*	*KHA*	*MAR*	*VAR*
18	15	17	1	1	2

These data support at least the *possibility* that villages whose names end with the suffix **-vali** were settled by Dravidian speakers, though the evidence is hardly overwhelming. The presence of hybrid village names, with an Indo-Aryan first part followed by a Dravidian suffix (such as **bāmaṇ-oli**, **dahi-vli**, **ānand-ośi**) is not an impediment to this assumption, given that similar hybrids (such as **cincoli najik**, **ali-vāḍā**, **āḷte kasbā**, with Perso-Arabic elements, and **cun kheḍi phoresṭ** (forest), **āmbulagā men** (main), **mālegāv kaimp** (camp) with English elements) have been formed in more recent times. On the contrary, such hybrids probably attest to a period of extensive bilingualism, probably involving both Indo-Aryan and Dravidian speakers.

9.4. Conclusions: Dravidian languages in Maharashtra

The differences in spatial distribution between the suffixes **-vali** and **-gāv** suggest that the early population of Dravidian speakers was probably concentrated in the Konkan and the southwestern part of the plateau region (Figure 9.5), whereas the incoming Indo-Aryan speakers were initially concentrated in the central and eastern

parts of Maharashtra (see Figures 9.3 and 9.4).[25] It is not possible to be certain whether the relatively few cases of names in **-vali** and other Dravidian place-name elements in the eastern regions are the residue of pre-Indo-Aryan Dravidian-speaking settlements, or whether they represent later migrations from the west. One point in favor of the former solution is the presence of Dravidian-derived river names in this area, as pointed out by Witzel (1999b: 73).[26] The name of the easternmost region, Varhad, is clearly non-Indo-Aryan, and may be derived from a Dravidian or Munda source.[27] Thus there is a good possibility that the whole of Maharashtra was occupied by Dravidian speakers at the time the first speakers of Indo-Aryan entered the area. In either case, it is probable that speakers of Dravidian maintained their language longer in the western and southern parts of the region. An alternative possibility is that there were two separate movements of Dravidian speakers into Maharashtra: one which covered the whole region, perhaps rather sparsely, and a later one which was concentrated in the Konkan. (See further discussion later in this section.) The concentration of village names with the suffix **-vali** leads to the inference that many of the Dravidian-speaking occupants of Maharashtra were probably associated with maritime occupations such as fishing or sea trade.

Any attempt to link the linguistic situation with the known archaeology of the region must look at the archaeological connections between the Deccan and neighboring regions. The Allchins state that during the period 2500–2000 BCE, when the first agricultural settlements were developing in the northern Deccan, the culture of this region showed close links with areas to the north, particularly toward the end of this period (Allchin and Allchin 1982: 352). On the other hand, S. B. Deo (as noted above) has mentioned pre-Chalcolithic Neolithic elements in the northern and Central Deccan which suggest links with the southern Neolithic cultures of the Andhra-Karnatak region (Deo 1982: 17–18). The Malwa Culture flourished first in central India (2300 ± 70 to 1445 ± 130 BCE at Navdatoli) and was extended later into the Deccan (1565 ± 95 to 1025 ± 170 BCE at Inamgao). The Jorwe Culture, which succeeded the Malwa Culture in a number of sites, existed during the period 1400–700 BCE on the basis of C^{14} measurements (Deo 1982: 23–4).

It is tempting to link the appearance of Dravidian place names in Maharashtra with the earlier, pre-Malwa period (3000–2500 BCE) – the same timespan assumed for Proto-Dravidian (see **8.1**, **8.20**) – given its connections in material culture to southern Neolithic sites like Utnur, Piklihal, Tekkalakota, Palavoy, Paiyampalli, Sanganakallu and Hallur. If we then assume, with the Allchins, that the Jorwe culture (which Deo has dated to *c.*1565 ± 110 to 1290 ± 95 BCE) represents an intrusion of Indo-Aryan speakers from the north, what then of the intervening Malwa Culture with its links to central India? The most likely assumption, if we are required to make a choice, is that speakers of Austro-Asiatic language(s) were prominent there, though there is nothing to deny the possibility that it was a multilingual culture, including Dravidian, Austro-Asiatic, and possibly other languages such as the ancestor of Nahali (see **3.2**).[28]

If there were speakers of a Dravidian language in Maharashtra in the early or mid-third millennium BCE, they would have been speakers of Proto-Dravidian or

pre-Proto-Dravidian (see **8.41**). The expansion of Proto-South Dravidian, which probably began in the early second millennium BCE (**8.42**), could have been responsible for the frequent occurrence of Dravidian place names in the Konkan and the western Deccan, as well as in Gujarat and Sindh (see **9.5**). Thus, as suggested earlier, it may be necessary to assume two prehistoric "invasions" of Maharashtra by Dravidian speakers. A thorough study of the Dravidian lexical elements in Marathi–Konkani, to be undertaken soon, may be able to help resolve these questions.

The mid-second millennium BCE, the presumed beginning date of the Jorwe culture, fits with the earlier suggested dating for the movement of outer-group IA speakers into the Indian peninsula (see **6.42(1)**). Given the extent to which these speakers appear to have been influenced by Dravidian languages, a symbiotic relationship with Dravidian-speaking groups is not difficult to imagine. This would imply a society in which languages of both families coexisted, one language perhaps functioning primarily as a domestic and ingroup language while the other served that purpose for some groups, and that of an intergroup *lingua franca* for the larger society. The widespread occurrence of Dravidian place names would suggest that a Dravidian language served the latter purpose at some period in parts of Maharashtra.

The shift in Maharashtra from predominantly Dravidian-speaking to predominantly Indo-Aryan-speaking was probably a long-drawn-out, gradual process, the first stage involving small groups of settlers who used an Indo-Aryan language for ingroup communication and a Dravidian language for intragroup interaction. The final stage may have involved high-caste Indo-Aryan speakers who obtained rights to land in the area through their kinship and political links to northern kingdoms. The arrival of this class of people may have led to the creation of the three-tiered structure – landlords, tenant-overseers, and laborers – which Southworth (1971) viewed as an essential part of the process of the "Dravidianization" of Marathi.

The area identified here as the main region of Dravidian habitation, that is, Konkan and the western Deccan, is outside the area of the archaeological sites discussed earlier. In fact, southern Maharashtra-desh is geographically intermediate between the Deccan sites of Inamgav, Chandoli, Sonegav etc. and the Southern Neolithic sites on the Krishna River (Utnur, Piklihal, etc.). Thus it is not unreasonable to assume that Dravidian forms of speech survived longer in this area than elsewhere in Maharashtra. The Konkan region has until modern times always been somewhat isolated from the plateau, both by its geography and its ecology. The speech varieties known collectively as "Konkani" cover a wide range, some relatively close to the standard Marathi of the Desh, and others quite different from it. All of these varieties are ultimately related to standard Marathi, but divergent from it in ways which indicate centuries of attenuated communication (see Southworth 1976b). This divergence probably began in the period when speakers of Marathi began to settle in the Konkan, probably after the period of the earliest Old Marathi in the tenth century CE. Thus it is possible that Dravidian languages might have survived there even into the second millenium CE.

9.5. Beyond Maharashtra

(*A*) *Gujarat.* H. D. Sankalia (1949) has indicated the presence of place names with the suffix **-palli** in inscriptions in Gujarat, as well as names ending in **-oli** in modern Gujarati. His map 2 (1949: facing p. 80) shows names in **-oli** located in eastern Gujarat, extending from the coastal area north of Maharashtra northward as far as Patan, as well as in southern and coastal Saurashtra.

(*B*) *The Indus Valley.* A brief investigation of the place names contained in the *Gazetteer of Pakistan* (Garren and Page 1983) reveals the occurrence of place names ending in **-wāli** (as written in the *Gazetteer*) in the northern Indus region, that is, Pakistani Panjab. Somewhat unexpectedly, this suffix is quite rare in Sindh. However, Sindh shows names in **-wāri**/**wari** (as written in the *Gazetteer*), appearing in similar contexts to those in which **wāli** occurs in Panjab.[29]

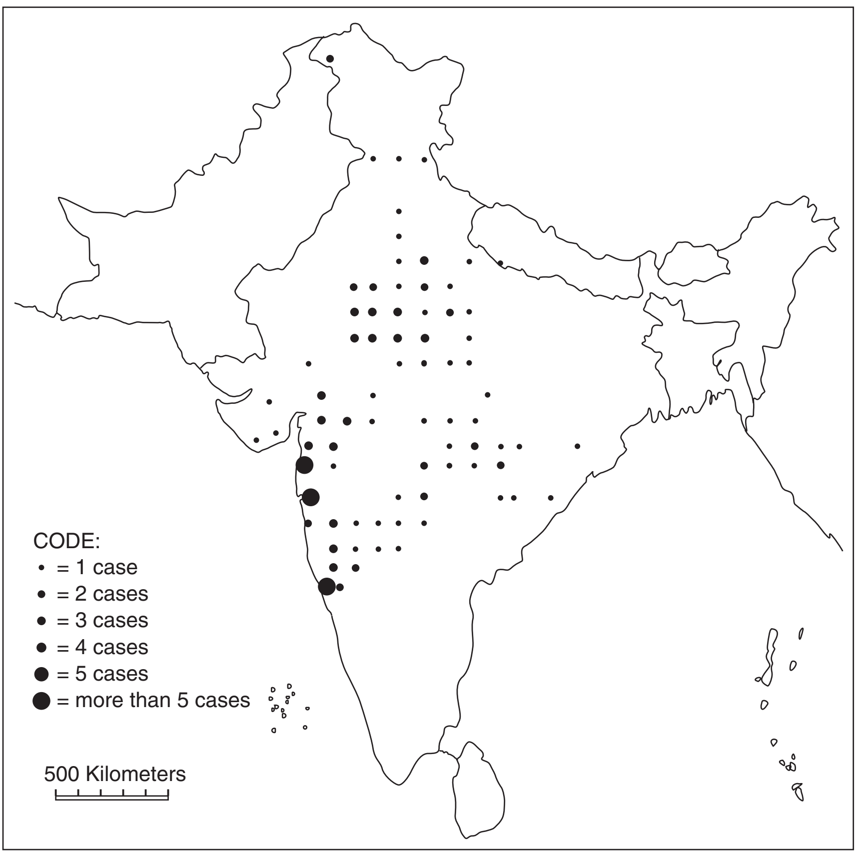


Figure 9.18 Distribution of **-v(a)li** and related suffixes in India.

Source: US Board on Geographic Names 1952.

(*C*) *Elsewhere in South Asia.* Figure 9.18 shows the distribution of **-vali** and related suffixes (**-vli**, **-oli**, **-olī**) in India, based on the US Government's list of official standard names for India (US Office of Geography 1952). The locations are approximate, as the coordinates have been rounded to the nearest whole degree of latitude and longitude (e.g. 18′ 16″ N to 73′ 46″ E has been converted to 18′ N to 73′ E).[30]

Figure 9.18 indicates that in India as a whole, this suffix shows a continuous distribution: while **-paḷḷi** (Kannada **-haḷḷi**) occurs throughout the southern Dravidian-speaking states, **-vali** and variants are found through western India (Maharashtra, Gujarat, Rajasthan) to Panjab, and into the upper Ganga Valley (Haryana and western Uttar Pradesh). This map does not show cases in Sindh and Pakistani Panjab (**B** above). These data show a few cases in central India, but none in the eastern states except for Orissa, which borders on Andhra Pradesh and has been an area of contact with Dravidian languages for many centuries.[31]

The distribution of this group of suffixes can be usefully compared with that of another suffix: **koṭ** (OIA **koṭṭa-**, PD ***kōṭṭ-**). This word, which means 'fort' in Indo–Aryan as well as Dravidian (probably derived from an earlier meaning 'wall'), has generally been regarded as of Dravidian origin (see **3.22A(4)**), though it is attested in one of the Ashokan inscriptions in the mid-third century BCE,[32] and can thus be regarded as a MIA word from at least that point onward. Parpola (2006b: 286) points out that names with this suffix occur primarily in two parts of South Asia: in the extreme northwest, and in the Dravidian south, where it occurs as a place-name suffix in the four southern states (George 1986: 147–9). This item was not included in the discussion of Maharashtra since the suffix **-koṭ** occurs only twice in the *Gramasuci*, and these occurrences are in the eastern part of the state. A 10 percent random sample of names ending in this suffix in India as a whole shows the following distribution:

	As suffix[33]	*As initial element*[34]
Kerala	1	5
Tamilnadu	6	6
Karnataka	2	8
Andhra Pradesh	5	15
Maharashtra	1	1
Gujarat	1	6
Rajasthan	1	12
Madhya Pradesh	5	18
Bihar		1
West Bengal		1
Orissa		6
Panjab	3	5
Himachal Pradesh		4

Source: US Office of Geography 1952.

This distribution of the suffix forms is similar to that of **-vali**, the main difference being that Uttar Pradesh and Haryana are not represented; the five occurrences in Madhya Pradesh are all in the southern part of the state in areas adjacent to Maharashtra or Andhra Pradesh. As initial element, the distribution is somewhat wider, the main difference being the larger number of cases in Madhya Pradesh.

A 10 percent sample of place names in Pakistan ending in **-kot** or **Kot** yields a total of 18 of which 9 are in Panjab, 1 is in Sindh, 1 is in Baluchistan, and 5 are in other areas. Of 36 names beginning with **Kot** in the sample, 31 are in Panjab, 1 is in Baluchistan, and 4 are in other areas, with none in Sindh.

In South Asia as a whole then, these two suffixes show a continuous distribution, including the southern states (in which they clearly have ancient origins), western India, the Indus Valley, and the northwest of India, with some cases in Orissa bordering Andhra Pradesh. The somewhat wider distribution of **-koṭ** toward the east of India may be the result of its having been borrowed into colloquial Indo-Aryan in (or before) the late MIA period.[35]

9.6. Conclusions

The primary evidence for the existence of Dravidian place names in Maharashtra consists of an estimated 800 or more village names with the suffixes **-v(a)li** or **-oli**, probably derived from the Dravidian **paḷḷi**, a word meaning 'hamlet' or 'village' in Proto-Dravidian which occurs widely as a place-name suffix in South India (see **9.21**). The differences in distribution between the suffixes **-vali/-oli** and **-gāv** (← OIA **grāma**, a word of Indo-European origin) suggest that different parts of Maharashtra were probably settled by Dravidian and Indo-Aryan speakers, with the Dravidian areas mainly in the Konkan (coastal region) and the southeastern part of the Deccan plateau; see Figures 9.3–9.6. Apart from the fact that the regions in which these suffixes are found are contiguous with areas to the south which show the Dravidian forms of the same suffixes, the probability of Dravidian influence is supported by other place-name suffixes, as well as some initial place name elements, in those same parts of Maharashtra (**9.24**, **9.3**). At the same time, there is some uncertainty because of the fact that the majority of village names in Maharashtra are without certain etymologies.

Place names with the suffix **-v(a)li** and variants also occur in Gujarat, Sindh, East and West Panjab, Haryana, western Uttar Pradesh, and Orissa (though not elsewhere in eastern India or Bangladesh), albeit with less frequency than in Maharashtra. Figure 9.18 shows that the whole region in which these suffixes are distributed is contiguous with Maharashtra and/or the Dravidian south. Thus these regions may also have been home to speakers of Dravidian languages in the past, though it has not been possible here to investigate this wider region in the same detail as for Maharashtra.

The presence of certain phonological changes (such as the change **-p-** → **-v-** in **-vali** ← **paḷḷi**) implies that the village names containing this suffix must date

at least to the early first millennium CE, if not earlier. Since there is no linguistic evidence which would fix the earliest possible date of the appearance of these place names in Maharashtra, there is no bar to assuming that they existed much earlier – say, as early as the second millenium BCE. The existence of Dravidian river names in northeastern Maharashtra (see **9.4**) suggests that all of Maharashtra might have been Dravidian-speaking at an earlier time, perhaps as early as the Proto-Dravidian period (third millennium BCE). Even if so, it is also possible that the widespread occurrence of Dravidian place names in coastal and southwestern Maharashtra, as well as in Gujarat and elsewhere, resulted from a later movement of Dravidian speakers from South India. Whatever may be the case, we are probably justified in concluding that by the early first millennium CE extensive areas of Gujarat, coastal Maharashtra, and southern Maharashtra were occupied by a population that used a Dravidian language for daily interaction – either as a primary home language and/or a *lingua franca*.

9.7. Summary

This chapter discusses eight Marathi place-name suffixes of probable Dravidian origin, and finds that their distributional patterns agree in that they occur with greatest frequency in the Konkan (coastal Maharashtra) and in the southwestern part of the Deccan plateau, an area in which suffixes of Indo-Aryan origin (such as **-gāv** ← OIA **grāma**) are of relatively lower frequency (see Figures 9.3, 9.4). It can be concluded therefore that these areas were probably inhabited by Dravidian-speaking people earlier, possibly as early as the mid-third millenium BCE, corresponding to the "pre-Malwa Neolithic" phase in Maharashtra which has links in material culture with the southern Neolithic (**9.4**). Going beyond Maharashtra, two suffixes of Dravidian origin are found in a continuous distribution from the Dravidian south, through the coastal area of western India (Maharashtra and Gujarat), into the Indus Valley and northwest India, suggesting that these areas may also have been home to Dravidian speakers earlier (Sections **9.5**, **9.6**).

Notes

1 Linguistic features attributable to Dravidian include

(1) a shift in the system of verbal negation from the older Indo-Aryan system of prefixing a **na-** to the verb (as in Skt **nāsti** 'is not', i.e. **na** 'not' + **asti** 'is') to a system of negative auxiliary verbs, used for different kinds of negation (e.g. **karu nako** 'don't do', **karu naye** 'shouldn't do', **nāhi** 'is not', **naste** 'is not' [habitually], etc.) – which is parallel, and identical in many details, to the system found in several Dravidian languages. This change is first found in texts a century or so *after* the appearance of Old Marathi in the eleventh century CE, indicating the probable presence of substantial Marathi–Dravidian bilingualism in the area as late as the twelfth century CE (see Southworth 1976b);

(2) an erosion of the inherited Indo-Aryan syntactic ergative construction in the direction of the Dravidian pattern of subject agreement (this change also appears to have taken place after the Old Marathi period, ibid.);

(3) a use of relative verbal participles which is more like Dravidian than other Indo-Aryan languages, and which is not inherited from earlier Indo-Aryan (see Nadkarni 1975);
(4) the innovation of a distinction between inclusive and exclusive first-person plural pronouns, found generally in Dravidian but not elsewhere in Indo-Aryan except in the western region, that is, in Marathi, Gujarati, and Sindhi, and in some Rajasthan dialects (see Southworth 1971);
(5) Words of Dravidian origin in Marathi basic vocabulary such as **ḍoke** 'head', **lek** 'boy, son', **leki** 'daughter', **-kaḍe** 'toward', as in **gharā-kaḍe** 'toward the house, homeward';
(6) phonological: a shift of stress accent to the initial syllable of words (as in Dravidian, and as opposed to its position in northern Indo-Aryan; see **3.31**, **5.22C**).

Other evidence derives from the work of Trautmann, a historian, who has shown that features of "Dravidian" kinship systems (particularly cross-cousin marriage and terminological categories compatible with it) are found in the southern part of the present Indo-Aryan-speaking zone (Trautmann 1981). In addition, the ancient South Asian historical tradition lists Maharashtra as one of the "Dravidian countries", though this is often dismissed by Marathi speakers as North Indian chauvinism. (See the final maps in Schwartzberg 1978, which show the Panca-Dravidas as 'Andhra, Dravida, Karnata, Maharastra, and Gurjara', and the Panca-Gauda as 'Sarasvata, Kanjakubja, Gauda, Maithila, and Utkala'. The medieval Yadava (OIA **yādava**) kingdoms of Maharashtra are linked to the traditions of the puranic Yadavas, who were often considered to be Dravidian or Dravidian-influenced (see **6.3** and Thapar 1975). The historian S. B. Joshi claims that Maharashtra was originally a Dravidian-speaking area, and that even as late as the 12th century CE the Yadavas of Hoysal identified themselves as speakers of Kannada while the Yadavas of Devagiri officially supported Marathi (Joshi 1951, 1952, quoted in Deshpande 1979: 102).

2 The data used for this chapter, from Apte's *Gramasuci* (Apte 1967), were based on the earlier district organization, shown schematically in Figure 9.2A. The new districts created by district reorganization, indicated by a distinctive font in Figure 9.1, are: Raigad (formerly Thane), Sindudurg (formerly part of Ratnagiri), Jalna (formerly part of Aurangabad), Latur (formerly part of Usmanabad), and Garchiroli (formerly part of Chanda, now renamed Chandrapur).

3 The total state population used in calculating the percentage for each district was 69 million; this excludes Greater Bombay, with 9.9 million.

4 Khandeshi, known locally as Ahirani, "has suffered the fate of being regarded as a dialect of Marathi, of Gujarati, of Bhili, of Rajasthani, and as an independent language" (Masica 1991: 453), and is perhaps best treated as a non-literary language which is transitional between Gujarati and Marathi, being mutually unintelligible with both (op. cit. 17).

5 V. N. Misra (2001: 516) gives a dating of 1700–1450 BCE for the Malwa culture at Navdatoli; see Figure 10.1.

6 Misra (2001:517) dates Early Jorwe at 1500–1200 BCE, and Late Jorwe at 1200–900 BCE; see Figure 10.1.

7 An inscription of 928 CE in Sanjan (Konkan) refers to a **pañca-gauḍa-mahāparṣad** (presumably 'Council of Five Gauda (group)s'). This is the area where even today the Gauda Saraswat Brahmans reside (Banerjea 1960: 102). I am grateful to Michael Witzel for this reference.

8 For example, **vāḍi** and **pāli**, which are common suffixes, also occasionally appear as names.

9 In Marathi the vowels **a i u** → **∅** (i.e. are lost) in the context VC_CV, therefore the suffix **-vali** appears as **-vli** after a vowel, as in **bori-vli**, **ambi-vli**.

10 Regarding the phonology of these forms,

(1) as noted above, **p** → **v** intervocalically in Marathi, as in most Indo-Aryan languages, before the mid-first millennium CE; thus an older ***nāpaḷḷi** would end up as ***nāvaḷḷi** after this change;
(2) $VC_1C_1 \rightarrow VC_1$, with lengthening of the vowel if stressed: thus ***nāvaḷḷi** would end up as **nāvaḷi**, whereas **paḷḷi** as a single word (or as the first part of a name) would be stressed on the first syllable, which would be lengthened, yielding **pāli**.

11 The following occur: **pāl**, **pālakhā**, **pālakhi**, **pālam**, **pālaye**, **pālā**, **pāle**, **pālavṭi**, **pāloḍ**, **pāloḍi**, **pālorā**, **pālhi**, **pālhe**, **pālheri**.
12 In the early Tamil literature, **paṭṭiṇam** refers to towns, particularly port towns; for example, the Chola capital at the mouth of the River Kaveri was called **Kāvēri paṭṭiṇam** (Nilakanta Sastri 1955: 22).
13 See notes 9 and 10 for the phonology of these forms.
14 Note that villages in Maharashtra, especially small hamlets of agriculturalists, are not infrequently walled.
15 There are also cities or towns named **Pāṭan** or **Pāṭaṇ** in Maharashtra, Gujarat, and Nepal, and possibly elsewhere.
16 Stampe's Munda materials (Stampe 2003) contain two items which are of possible relevance here:

(1) Sa. **ḍi** 'settlement, village'; Mu. **ḍi** 'settlement, village'; KW **ḍi** 'village';
(2) Sa. **atɔ** 'village, hamlet'; Mu. **hatu ~ atu** 'village, hamlet'; Ho **hatu** 'village, hamlet'; Bj. **hathu** 'village, hamlet'; As. **vatu** 'village, hamlet'; KW. **vatu** 'village, hamlet'.

17 The suffix **ner** is the Marathi reflex of OIA **nagara**, which first occurs in OIA relatively late (late Vedic) and is probably ultimately of Dravidian origin (see **3.22A(3)**).
18 Molesworth's Marathi–English dictionary (Molesworth 1975) lists **vaśi** 'name of a saltwater fish'.
19 Those of presumed Indo-Aryan origin include **šeṇ** 'dung' (OIA **šakana**), **āmbā** 'mango (tree)' (OIA **āmra**)', **bor** 'jujube' (OIA **badara**), **cinc** 'tamarind (tree)' (OIA **cincā**), **vaḍ** 'banyan tree' (OIA **vaṭa-**), **bel** '*Aegle marmelos*'(OIA **bailva**), **māng** (a low caste) (OIA **mātanga**), **koḷ(i)** 'Koli' (a caste of coastal fishermen) (OIA **kola**), **ḍomb** 'Domb' (a low caste) (OIA **ḍomba**), **cor** 'thief' (OIA **cora**), **dahi** 'curd' (OIA **dadhi**), **ṭek** 'hill' (OIA ***ṭekka**), **ḍer** 'resting place' (OIA **ḍera**), **bhāl** 'spear' (OIA **bhalla**), **budhā** 'bottom' (OIA **bundha**), **cāph** (OIA ***carpa** 'flat'), **gāḍh** (cf. **gāḍhav** 'ass' ← OIA **gardabha**), **ghoḍ** (cf. **ghoḍā** 'horse' ← OIA **ghoṭaka**), **goṭh** (cf. **goṭhaṇ** 'cowpen' ← OIA **go-sthāna**), **jāmba** 'rose-apple' (OIA **jambu**), **kaṭ** '*Saccharum sara*' (OIA **kaṭa**), **kāṭ** 'thorn' (OIA **kaṇṭa**). However, it should be noted that Dravidian origin has been suggested for many of these.
20 A number of these names contain the vowel **i**, which can be regarded either as a suffix on the specific, or as some sort of connector between generic and specific. This is noticed most frequently in the names for social groups such as **ḍomb-i-vli**, **koṇḍ-i-vli**. This **i** does not appear to be motivated from Marathi; while there are some names of this type which end in **-i** (such as **koḷi** 'person of Koli tribe', **māḷi** 'gardener'), those without this ending never add it in compounding. A possible source for this **i** would be the Dravidian ending **-(i)n** which marks "adnominal relationships" in several SDr. and CDr. languages (Zvelebil 1977: 31).
21 Some additional items may be mentioned, although they fail one of the tests for inclusion in the given table. These include three with acceptable Dravidian etymologies, which however fail to satisfy the distributional requirement: (1) **ār** (DEDR 4233 **āru** 'river' or 315 **ār(am)** 'mountain ebony'); (2) **kāṭ/kāḍ** (DEDR 1206 **kāṭu** 'forest'). The following appear in the required distribution, and thus *may* be of Dravidian origin, but no satisfactory etymology has been found for them: (3) **āṭi/āḍi** (?DEDR 63 **āṭi** 'foot, base, bottom' – poss. 'foot/base of mountain'?); (4) **āḍ**.

22 PSD ***ampi** means 'boat' (DEDR 177); DEDR also mentions a Marathi **ambi**, **ambekari** 'a boatman (about Kolhapur and in Carnatic)'.

23 See **3.22A(4)**.

24 Cf. also Koṇḍa, a Dravidian language of Central India (see Krishnamurti 1969); these ethnic names may also be ultimately derived from PD ***ku*nr*a-** 'mountain' (DEDR 1864).

25 It may be noted here that Vidarbha was the first settled kingdom south of the Vindhyas to be mentioned in Sanskrit literature (Nilakanta Sastri 1976: 69).

26 "…all around vidarbha, the first Drav. river names are met with : the *pūrṇā* (< **pēṇ*) west of it, the *vēn-gaŋgā* east of it, and the *pain-gaŋgā* south of it. They all are adaptations of a Drav. term for rivers, DEDR 4160a **pēṇ-: *pēṇ-V-* 'to twine, twist'. It seems that the area which still has a Munda name in the Vedic middle period (*vidarbha*) has also received a Dravidian overlay."

27 The name **varhāḍ** can plausibly be derived from a Dravidian ***vaṭa-kāṭu** 'northern forest' (DEDR 5218 SD ***vaṭa** 'north' + DEDR 1438 PD ***kāṭu** 'forest'). Witzel sees an Austro-Asiatic prefix **vi-** in the word **Vidarbha**, as in other names such as **Vindhya** (mountains) and **Vibhindu** (name of a people) (2001: 73).

28 The search for traces of Austro-Asiatic and other languages in Maharashtra cannot be dealt with within the scope of this chapter, but will be the subject of a subsequent investigation. Apart from the presence of the Korku (Munda) and Nahali (isolated) languages in the region, and a few hints like the name Vidarbha (see above), there are probably many clues to be found in Maharashtrian toponyms and even personal names (e.g. caste and family names). Even in the general vocabulary of Marathi there are numerous words needing further investigation, such as **ḍoŋgar** 'hill, mountain' (a basic vocabulary word), which is widespread in Indo-Aryan (cf. Panjabi **ḍogrī** 'mountain-dweller, mountaineers' language') and occurs in some Munda languages (e.g. Kharia **ṭoŋgrī** 'hill, small jungle') but does not appear to be Proto-Munda. Such words may provide clues to the earliest inhabitants of the region.

29 It is not clear whether the **R** of Garren and Page 1983 represents Sindhi **r** or **ṛ**. If **r**, it would be expected to go back to a single **l** in OIA (cf. Sindhi **pharu** 'fruit' ← OIA **phala**, **tharu** 'dry land' ← OIA **sthala**. OIA and MIA **ll** generally appear in Sindhi as **l**, as in **phulu** 'flower' ← OIA **phulla-**, **thalu** 'firm land' ← OIA **sthalya**, MIA **thalla**). If the graphic **R** represents Sindhi **ṛ**, then it would most likely go back to an earlier **ḍ(ḍ)** or **ṭ(ṭ)**. In either case, it is not the expected result from an OIA **ll** ← Drav. **ḷḷ**. However, the parallels in the uses of these two suffixes (**wāli** in Panjab and **wāri** in Sindh) should be taken into account:

Khadwāri	Khānwāli
Goth Allahwārio	Jakhio Dhok Gangānwāli
Chhānwari	Chandarwāli
Panwāri	Panwāl
Phulliānwāri Nai (river)	Phulānwāli, Chhāpiānwāli
Wekhewāri Dhora (stream)	Badarwali Kot (fort)

Another possibility is that Sindhi **-wāri** may represent Dravidian **pāṭi** or **paṭṭi** (see **9.23**).

The Panjabi suffix written **-wāli** in the data may be different from the ubiquitous **-vālā**/**-vāli** suffix of Hindi–Urdu and Panjabi found for example in Hindi **gharvālā** 'householder, member of a household', which is probably derivable from OIA **pāla(ka)**, as seen for example in **dvāra-pālaka** 'door-keeper' (thanks to Dr. Surendra Gambhir for this suggestion).

30 The reliability of these data may be questionable, inasmuch as the total number of names shown in Maharashtra is less than one-tenth of the number shown in Figures 9.3 and 9.6. On the other hand, this map agrees with those figures in showing the majority of Maharashtrian cases on or near the coast. Thus it may be assumed until proven

otherwise that there is no bias in the data apart from that of size – it is probable that only villages above a certain size have been included.

31 The two cases in Kashmir (Jundoli and Mir Wali) are rather isolated, and these suffixes may possibly be of different origin. Another group of suffixes, **-auli/-aulī/-āuli**, appears mainly in central and eastern Uttar Pradesh and Bihar (58 cases), overlapping only slightly with the **-vali** group, along with a few (7 cases) in Maharashtra. Inasmuch as these distributions are separate, the UP cases may have a different origin. The suffix **-pāli** (see Figure 9.10 and discussion there) shows a different distribution, with 9 cases scattered through Maharashtra and a separate group of 13 in Orissa and neighboring Jharkhand.

32 The word in question is **koṭṭaviṣesu**, which occurs in a short inscription at Sarnath (near Varanasi) and is glossed as 'domaines à forteresses' or 'territoires fortifiés' in Bloch (1950: 153).

33 (Based on 10 percent sample.) Suffix forms occurring in the sample are: **kot** (3), **kota** (3), **kottai** (2), and **cottah**, **koti**, **kote**, **gota** (1 each).

34 Source: US Office of Geography 1952: 445–9. Forms include **kot**, **kota**, **kott**, and **kotta** (the last two mainly in the southern states).

35 Though he acknowledges the presence of the word ***koṭṭ-** in Proto-Dravidian, Parpola believes that it may ultimately be of Indo–Aryan origin (2002a: 286). The Dravidian semantic development ('wall' → 'fort') would argue against this, but in any case, borrowing between OIA and Dravidian in either direction would imply the presence of Dravidian speakers in the northwest of the subcontinent at an early period. The only alternatives are accidental resemblance or borrowing from a third source; the "Indus" languages could be such a source, though the relative lack of the suffix in Sindh (see later) raises some questions.

10

HISTORICAL LINGUISTICS AND ARCHAEOLOGY IN SOUTH ASIA

Contents

10.1. Introduction

A central motivation for the writing of this book is the conviction that the conclusions of linguistic archaeology need to be tested against reality: however rigorous our approach may be, at some point we must attempt to relate our findings to the hard facts – the physical remains, times, and places which archaeologists and other prehistorians deal with. At the very least, we need to learn how far it is possible to make such correlations. Some of the problems involved in the comparison of linguistic and archaeological findings have been discussed in **1.7**. The intervening chapters (particularly Chapters 3 and 5–9) contain various inferences about prehistoric events based on linguistic data. These inferences are based in the first instance exclusively on linguistic evidence; only after the linguistic inferences have been spelled out, has there been any attempt to look for correlations in the archaeological or historical literature. The following Section, **10.2**, presents a brief summary, roughly chronological, of the linguistic prehistory of South Asia based on the inferences made in those chapters. In the final Section, **10.3**, I return to the question of the relationship between linguistic and archaeological evidence, and its implications for the future of communication between the two disciplines.

10.2. Summary of South Asian linguistic prehistory

This section summarizes the conclusions of Chapters 3 and 5–9, along with previous research by the author and others, regarding the prehistoric linguistic situation in South Asia. The evidence for the locations of languages and linguistic features is linguistic and philological, based on the known locations of languages and linguistic features on the ground, including the evidence of ancient texts. The chronology, on the other hand, depends substantially on archaeological and other historic/prehistoric evidence: for example, the placing of Proto-Dravidian in the third millennium BCE is heavily dependent on its presumed identification with Southern Neolithic sites (**8.4**). Section **10.21** is a chronological overview, and **10.22** briefly looks at the linguistic prehistory of the major regions of the subcontinent.

10.21. Chronological overview

Figure 10.1 shows the relative chronology of a number of important South Asian archaeological sites in the period between the fourth millennium BCE and the first millennium CE, which are referred to in the following paragraphs.

10.21A. Before 2500 BCE

The situation before 2500 BCE is depicted in Figure 3.1. See **3.2** for descriptions of the languages shown in the figure.

10.21B. 2500–2000 BCE: the Indus Valley

This period sees the main development of the Indus Valley culture complex, also known as the Indus (or Harappan) Civilization. We have no direct knowledge of the language(s) used in this culture complex, in spite of considerable effort expended by many scholars from different nations on the decipherment of the Indus inscriptions (see **3.1** and Possehl 1996). Nevertheless, certain general assumptions can be made about the language(s) used there in terms of the *roles* they must necessarily have played. This huge area combined numerous rural and urban environments, and exhibited the signs of distinct local traditions along with the wider communication networks necessary for trade and the transport of goods. The Indus seal inscriptions, whatever may be their linguistic nature, are indicative of a system of wider communication among the far-flung sites of the Indus Valley cultural complex. In addition, there were interactions with peoples at a very different level of socioeconomic development, such as the groups who provided copper to the people of Lothal (Possehl and Kennedy 1979).

Thus there inevitably existed regional forms of speech, which may have been very distinct from one another, possibly even unrelated languages, while some sort of *lingua franca*, or perhaps more than one, must have served for wider

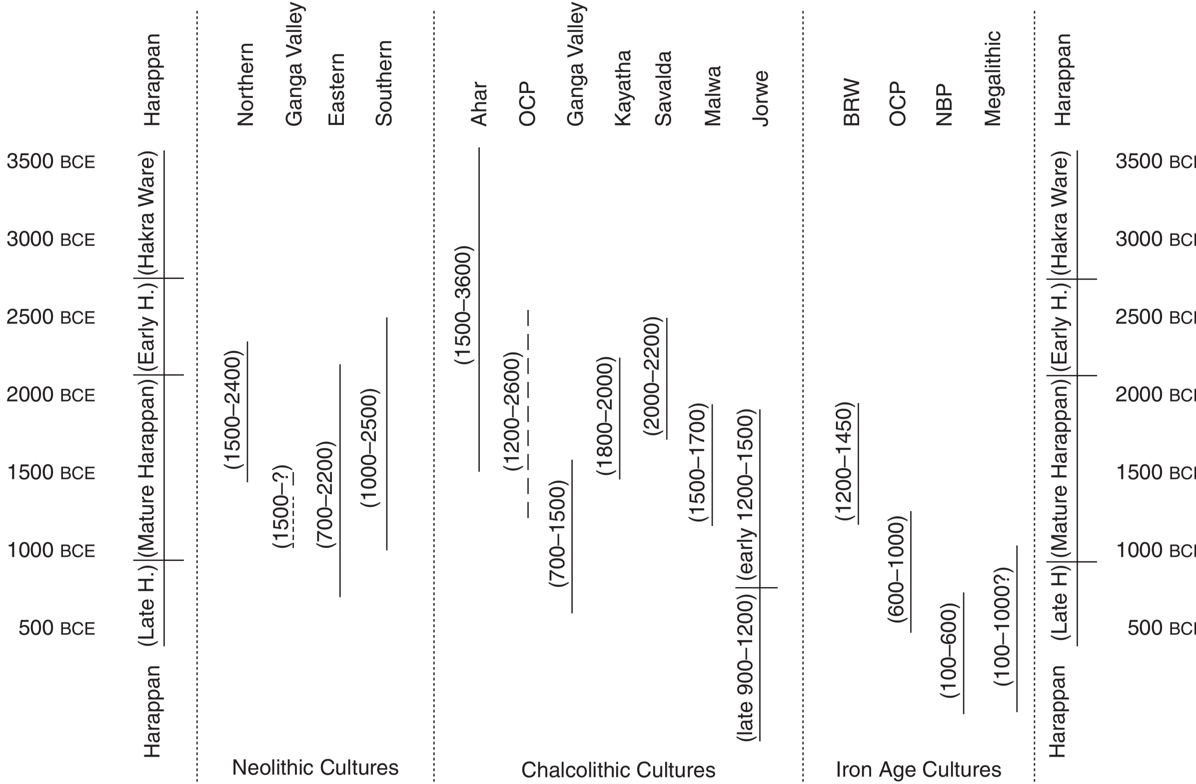


Figure 10.1 Relative chronology of South Asian archaeological cultures (Dates from V. N. Misra 2001).

communication. This is a pattern found in many parts of Asia at all periods of history, and was probably the situation in the ancient cultures of Mesopotamia.[1] In historic times in South Asia, in many areas there have existed three or even four levels of linguistic communication: (1) local village dialects and/or "tribal" languages, (2) dialects of wider use (such as Kanauji, Chattisgarhi Hindi, Tirunelvely Tamil, Nagpur Marathi, Telengana Telugu), sometimes based on the speech of a city or a large town, (3) regional languages (such as Hindi, Tamil, Oriya), and (4) interregional or national languages such as Sanskrit, Persian, and later English. Given the enormous geographical extent of the Indus Valley culture, it must be assumed that at least two or three such levels existed.

On the basis of current knowledge – which is admittedly fragmentary – the best candidate for the role of regional *lingua franca* in Panjab would be "Para-Munda", the name given by Kuiper and Witzel to the (presumably) Austro-Asiatic language which made numerous contributions to the early Rigvedic vocabulary (**3.21**). Given the existence of Munda languages in eastern India, which have proven affiliations with various languages of Southeast Asia, it must be assumed that languages belonging to this group were widely spoken throughout the northern part of the subcontinent, as indicated in Figure 3.1. In the Indus Valley, underlying Para-Munda, both chronologically and in terms of its sociolinguistic role, was the "Indus" speech, which must be assumed to have comprised many local dialects – if not distinct, possibly unrelated, languages. This "Indus" language (group) can be assumed to have existed in Sindh as well, in order to account for the similarities in words found in early Indo-Aryan and Dravidian (see **3.23**). One or more Dravidian languages may have served as a local language or as a *lingua franca* in Sindh, as well as in Saurashtra, though perhaps only at a later period. In Baluchistan we assume the existence of a language which has been dubbed "Meluhhan", which contributed words for Indus Valley products to the languages of Mesopotamia.

South India. Roughly contemporaneous with the early Indus Valley culture was the beginning phase of the Southern Neolithic archaeological complex. It has been argued in **8.4** that this complex was probably associated with the reconstructed language known as Proto-Dravidian. In fact, the Proto-Dravidian stage may have begun a few centuries earlier if it was originally located in the Godavari–Krishna area, as suggested in **8.4**. This dating is consistent with what one would expect from the linguistic development of Dravidian (see **2.42**). Thus Proto-Dravidian was probably contemporary with early Harappan culture. The words listed in **3.23** are shared by early Dravidian and Old Indo-Aryan, but cannot be shown to be original in either, and thus have been assumed to be from the hypothetical "Indus" language(s) of the Indus Valley. In addition, the evidence for incipient social stratification, private property, and commercial activity in PD suggests the possibility of contact with complex societies in the Indus Valley or western/central Asia, assuming that PD is too early to have had contact with the second urbanization in South Asia during the mid-first millennium BCE. Though

the term "Proto-Dravidian" is used for this South Indian speech community of the mid-third millennium BCE, it should be emphasized that this was not necessarily the earliest speech community to use a Dravidian or Dravidian-related language (see **10.21D**).

In subsequent periods the Southern Neolithic complex expanded from its core area (along the present Andhra–Karnataka border) in all directions. This expansion, which is not as yet precisely dated archaeologically, was probably accompanied by the linguistic differentiation of Dravidian into three major branches: South, Central, and North. South Dravidian, the largest branch, further divided into SD1 (Tamil, Kannada, etc.) and SD2 (Telugu, Gondi, etc.). Each of these branches and sub-branches ultimately subdivided further, resulting in the 25-odd Dravidian languages known to us today. The Central Dravidian languages are now found mainly in the northeastern part of the peninsula, an area which also includes one of the North Dravidian languages, Kudux; the other North Dravidian languages are found far to the north (Malto) and northwest (Brahui); see Figure 2.3. This linguistic fragmentation of Dravidian, like the geographical spread of the Southern Neolithic culture, was a continuous process. Though we can reconstruct the vocabulary of Proto-ND, Proto-CD, Proto-SD, as well as that of the sub-branches Proto-SD1 and SD2, we should not imagine that any of these represents a language in the usual sense of the word, that is, the speech system of a uniform speech community existing at a particular time and place. Even Proto-Dravidian itself is subject to this stricture, in the sense that it represents a composite of speech varieties spoken over a longish period of time and throughout a wide area. (See **8.4** for further details, as well as the theoretical discussion in **1.31**.)

10.21C. 2000–1500 BCE

This period sees the addition of Indo-Aryan languages to the previously existing South Asian language families. The oldest parts of the Rigveda can be dated to the mid-second millennium BCE (see **2.24**), and internal evidence suggests a location in the northern Panjab.[2] The outer Indo-Aryan languages, which are less directly linked to the OIA texts, probably entered the Panjab–Sindh region (modern Pakistan) several centuries earlier, on the basis of the following arguments:

(1) Linguistic features of the outer languages are found in Sindh, Gujarat, Maharashtra, the Konkan, southern Madhya Pradesh, and the eastern region (Orissa, Bihar, Bengal, and Assam). In the west, particularly in Sindh and Gujarat, these features have been overlaid by features derived from the inner languages Hindi–Panjabi (see **5.3**), whose histories are more closely associated with the Indo-Aryan of the older texts. Thus it appears likely that the variety of OIA which was the source of the outer languages was already present in Sindh by the time of the Rigveda.

(2) Many of the words of Dravidian origin found in OIA texts of the Vedic period (see **3.22A**, esp. (1)–(3), **3.23A**) appear to derive from traditions outside of the

> mainstream. These include the sources of the Atharvaveda, which according to Emenean (1954) may contain materials which are as old as, or older than, the Rigveda, though transmitted in a more evolved linguistic form. It is likely that speakers of outer languages, already more integrated into the local cultures, may have served as a conduit for such forms. An example is the word **Yādava** (name of a people), attested first in the Rigveda and later associated with heterodox clans living in the area of present-day Gujarat. This word can be plausibly derived from a Proto-Dravidian ***yāṭu-van̲** 'goat/sheep herder' (lit. "sheep-man"), suggesting that clans of shepherds speaking outer Indo-Aryan dialects adopted a Dravidian designation for themselves, which was then transmitted to speakers of inner Indo-Aryan in the form of a clan name. In subsequent periods the carriers of the outer forms of Indo-Aryan probably moved southeastward from Sindh through Gujarat and Maharashtra, and thence across central India to the east (see the following section).

The developments described in this section are still rather controversial among historical linguists. The hypothesis of inner–outer Indo-Aryan has yet to be accepted by Indo-Aryanists, though it may be pointed out that much new evidence for it is being published here for the first time (Chapters 5 and 6). As to the Dravidian element in early Indo-Aryan, there are perhaps still some scholars who would deny its existence, at least for the Rigveda, though many others, including Mayrhofer (1953b, 1986) have – sometimes grudgingly – accepted at least some Rigvedic words as (probably) Dravidian.

10.21D. 1500–1000 BCE

Some Sanskritists believe that the first Dravidian loanwords in the Rigveda are found around 1200 BCE (see esp. Witzel 1999b). Witzel suggests the possibility that these words came from Dravidian speakers in the area of Sindh, reflecting the geographical expansion of the Vedic society at this period. It is possible of course that speakers of Dravidian languages were present in Sindh even in the preceding period (see **10.21C**), but that these vocabulary items did not find their way into the OIA ritual language until this period. In fact, section **3.22** presents evidence of contact between early Dravidian and Proto-Indo-Iranian, that is, the parent stage of Indo-Aryan and Iranian, as well as additional evidence for contact with Indo-Aryan speakers in Rigvedic times. As some of this is new evidence which has not yet been examined by scholars in the field, it cannot be called decisive at present.

If the early Dravidian speakers entered the subcontinent from the northwest, as suggested by some (see **2.4**), then the Dravidian-speaking groups in Sindh at this period may have been speakers of pre-Proto-Dravidian languages[3] who remained in Sindh while others moved into the peninsula. Alternatively, they may have been members of the Proto-Dravidian community of South India (see above) who migrated northward into Sindh. It is also possible that there was ongoing traffic between the Southern Neolithic culture and the Indus Valley from the mid-third

millennium BCE onward. (Note Allchin and Allchin's discussion of the ancient route connecting the Panjab and the Ganga Valley with the south, 1982: 249.) The possibility of communication between the two areas by sea should also not be neglected (see **8.42**).

By the end of this period (mid-second millennium BCE) a number of archaeological sites in Maharashtra show affinities with sites further to the south, and the probability is that the southern sites are older than those in the Deccan (Fuller *et al*. 2001). In this case, then, it would seem that the presence of southern elements in the northern Deccan may represent a northward expansion of the Southern Neolithic. This northern movement may be the source of the many Dravidian features found in the Marathi language, as well as the Dravidian place names found throughout Maharashtra and Gujarat (Chapter 9).

The other major development in this period is the (presumed) movement of the outer Indo-Aryan languages from Sindh into the Deccan, and thence into the eastern region (see **6.42**). We have as yet no means of linking these movements to any archaeological phenomena, and thus the chronology must remain vague. Roughly, we may say that speakers of outer Indo-Aryan were probably present in Sindh in the early centuries of the second millennium BCE, from whence they presumably spread southeastward into the Deccan and thence across central India south of the Vindhya complex and through the Balasore gap into eastern India, reaching Bengal and Bihar well before the time of the Buddha, say by the first half of the first millennium BCE (see **6.42**). Thus, we may assume that this process went on for about 1,000 years.

10.22. Linguistic prehistory by region

In an effort to make the picture of South Asian linguistic prehistory as clear as possible, this section presents brief summaries of the linguistic situation in each major region of the subcontinent, based on the account given in **10.21**. Figure 10.2 provides a regional breakdown of changes in the linguistic inventories of (1) the Indus Valley (most of modern Pakistan), (2) the Ganga–Yamuna Doab (roughly the old Uttar Pradesh, before the recent creation of the state of Uttaranchal), (3) a central segment consisting of Gujarat, Maharashtra, and southern Madhya Pradesh (now a new state, Jharkhand), (4) the eastern region (pre-partition Bengal along with Orissa, Bihar, and the older state of Assam), and (5) peninsular India (the four Dravidian-speaking states of Andhra, Karnataka, Kerala, and Tamilnadu). The evidence for these prehistoric inferences has been presented in **10.21**, along with references and cross-references, and will not be further referenced here.

Indus Valley. The "Indus" language (group) is probably the oldest detectable linguistic stage, both in Panjab and Sindh. An unidentified western Austro-Asiatic language designated as "Para-Munda" probably functioned as a *lingua franca* of this area. Both of these language groups may have been present as early

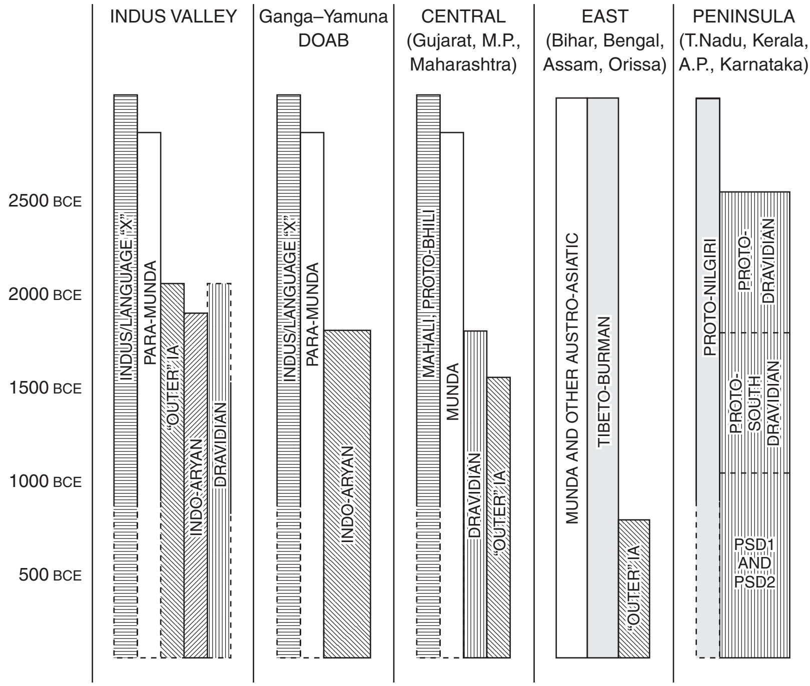


Figure 10.2 Chronology of South Asian linguistic prehistory.

as the seventh or eighth millennium BCE, when archaeologists record the beginnings of agriculture in areas bordering the Indus Valley. Indo-Aryan appears in the form of Vedic Sanskrit, dateable to the early second millennium BCE, probably preceded by outer Indo-Aryan. The arrival of Dravidian-speaking people in the area is difficult to date and is the subject of some controversy. Evidence of Dravidian borrowings in the Rigveda dates only from about 1200 BCE, according to some sources, yet it is nevertheless possible that Dravidian languages were spoken in Sindh and Saurashtra (and perhaps even in Panjab) considerably earlier. Ultimately, only Indo-Aryan survives in these areas, though Sindh shows traces of outer Indo-Aryan.

Ganga–Yamuna Doab. It is possible that the "Indus" language(s), along with Masica's Language "X", were spoken in this area as well as in the Indus Valley; in fact, these two terms may refer to the same language or language family (see **3.2**, item 3g). If so, it would probably predate the Para-Munda language(s) in this region. Indo-Aryan texts know of the Ganga by the late Rigvedic period, so it is likely that Indo-Aryan speakers reached this area by the early first millennium BCE.

Central. The isolated language Nahali, along with the inferred proto-Bhili, possibly represent the oldest detectable linguistic stratum in this region. Munda

languages were spoken at least in the eastern portion,[4] and Dravidian languages were present probably by 1000 BCE if not earlier. Dravidian place name suffixes are found in Maharashtra, Gujarat, and Sindh, and Dravidian may have played a role in the southern cities of the Indus Valley culture in Sindh and Gujarat. Outer Indo-Aryan languages probably appeared in this area by the mid-second millennium BCE or earlier.

East. Munda and other AA languages, along with Tibeto-Burman languages, are probably equally ancient in this region. Outer Indo-Aryan probably was spoken in North Bihar by the early first millennium BCE, possibly earlier in Orissa.

Peninsula. Proto-Nilgiri, an inferred language of the Nilgiri hills, is perhaps the earliest language for which evidence exists in peninsular India. If the identification with the Southern Neolithic archaeological complex is valid, Proto-Dravidian can be dated to the mid-third millennium BCE, or perhaps a few centuries earlier. Subsequent stages cannot be dated with much precision, though Krishnamurti discusses evidence for the split of Proto-South Dravidian into PSD1 and PSD2 around the end of the second millennium BCE (Krishnamurti 2003: §11.4).

10.3. Historical linguistics and archaeology in South Asia

The above-mentioned sketch contains a number of inferences about South Asian prehistory which are not visible from the archaeological record, though they do not conflict with it on the whole. The question is, to what extent can any of these assumptions or hypotheses be tested against the archaeological findings? As pointed out in **1.7**, linguistic events like language contact and geographical displacement imply various kinds of social change. The following paragraphs spell out the kinds of social changes implied by the specific inferences made in Chapters 3–9, and discuss the possible types of archaeological evidence that might be relevant to testing them.

Chapter 3 presents a picture of an ancient linguistic area, with different languages in contact with each other in various parts of the subcontinent, particularly in the Indus Valley (see **10.22**). Because of certain linguistic changes, particularly the adoption of the dental–retroflex contrast in consonants, it is inferred that the contact between the inferred "Indus" language and the Munda/Para-Munda languages was somewhat intense, implying a fairly high degree of socioeconomic integration (see **4.29**). The same was true later, of the contact between OIA (presumably both the inner and outer varieties) and the local languages, which presumably included both "Indus" and Para-Munda. Thus we infer some sort of economic interdependence in both of these cases. If "Indus" and Para-Munda were languages of the Indus Valley culture (respectively a local language and an interregional *lingua franca*), then it would not be surprising if such contact occurred; nor would it be surprising if early speakers of Indo-Aryan interacted with the local people in similar ways, given the need of pastoralists for agricultural produce. Interactions between Dravidian and Indo-Aryan speakers appear to be somewhat later, and perhaps occurred first in Sindh. Whether there are any

specific archaeological assemblages which could be linked with any of these contact situations would require a separate study.

Significantly, however, we are still left with the old mystery: how did the Indo-Aryan languages supplant those of the established Indus Valley culture? Even assuming, as is commonly done nowadays, that Indo-Aryan speakers only appeared in the Panjab after that culture had begun to decline and its communication and trade networks had been disrupted, this is still a significant question. If, as Witzel and Kuiper believe, Para-Munda was a major language of the Indus Valley culture – Witzel (1999b: 14) even suggests that it may have been the language of the Indus script – what happened to the local languages and their speakers, particularly the elites among them?[5] Linguistics cannot solve this problem, of course, and the linguistic perspective may even serve to emphasize its paradoxical nature. As noted in **1.7**, the social factors associated with language shift and language loss are usually economic – or more broadly, ecological: new languages are learned from ecological necessity, and existing languages are lost when they no longer fill ecological needs.

A possible suggestion from a sociolinguistic perspective is that an alternative communication network already existed before the breakdown of the older Indus network. This might have been the case if outer OIA-speaking pastoralists were already present in the Indus Valley before the decline of the Indus Valley culture (say by the early second millennium BCE) in sufficient numbers to have created their own interactional network within the framework of the Indus Valley society. This would presumably have been done by moving into niches which allowed them to establish relations with local agriculturalists, and perhaps also with town dwellers. There is evidence for pastoralism in the Indus Valley from the Early Harappan period on (Possehl 1999: 58ff.), and the possibility of Indo-Aryan speaking pastoralists in the southern Indus area has been mentioned before (**10.21C**). One would expect that such networks would be expanded as more and more OIA-speaking pastoralists entered the region. Such a hypothesis runs counter to what might be expected, since agriculturists tend to be more numerous than pastoralists in any area and thus the language of the latter is less likely to become dominant (Renfrew 1987: 271). However, if we are dealing with a situation with many local languages/dialects, the wider network of the pastoralists might become the more viable medium of regional communication, especially in a situation where the earlier network which people depended on was no longer functioning. Alternatively, Indo-Aryan-speaking pastoralists may have moved into an existing (?Para-Munda-speaking) network and gradually converted it to an Indo-Aryan-speaking network.

Chapters 5 and 6 deal with a problem in what historical linguists call subgrouping (see **1.22(B)**). A set of linguistic features which has a limited but continuous distribution among the NIA languages is used as a basis to posit two partly separate branches of Indo-Aryan – in the sense that the two legs of a pair of trousers are separate, each with its own (partly) separate history. From this linguistic history it is inferred that the two "legs" of Indo-Aryan traversed different

paths, one (the inner languages) from Panjab eastward into the Ganga–Yamuna Doab, the other (the outer languages) from Sindh southwestward through Saurashtra and Gujarat into the Deccan, and thence eastward to Orissa, Bihar, Bengal, and Assam, with subsequent westward movement from Bihar to Kosala (modern Avadh).[6] Such an inference raises several sets of questions: first of all, what conditions might have provoked these alleged movements? Were these simply cases of the "wave of advance" model (**1.72A**, Renfrew 1987: 124ff.), that is, were the Indo-Aryan speakers of these two groups mainly agriculturalists by this time, and moving as needed to find more land as the population grew? This might have been the case even if the overriding ideology of the society remained linked to pastoralism because it was the chosen activity of the elites. Were there differences in culture content between the two groups? The prevalence of goddess worship and snake worship have been reported in both Maharashtra and Bengal, though of course they are not unknown elsewhere.[7] In addition, the two great Sanskrit epics appear to belong respectively to these two branches of Indo-Aryan: the Mahabharata to the inner group, and the Ramayana to the outer (**6.31**). Another question concerns the inference that the two "legs" merged in the area of Avadh, where there is evidence of linguistic mixture. What might be the archaeological equivalent of this mixture? There also appear to be other transitional areas, for example Rajasthan and Bandelkhand (western Madhya Pradesh). Once these questions are asked, it is possible that linguists and archaeologists working together could find ways to test these hypotheses.

Chapter 7, a study of plant names and their histories, suggests several hypotheses, some of which reinforce those suggested by other evidence: for example, the sharing of a word for sesame between Mesopotamia (Akkadian **ellu**) and South India (SD1 **eḷḷu**) reinforces the hypothesis of the presence of Dravidian speakers in the prehistoric Indus Valley. A different kind of example is the suggestion that certain agricultural crops may have reached South India by sea from Southeast Asia (rice, ginger, sugarcane, coconut) or from Africa (sorghum).

Chapter 8 is a study of ancient (reconstructed) Dravidian vocabularies which leads to a number of inferences. An association is inferred between the reconstructed language known as Proto-Dravidian and the core area of the Southern Neolithic archaeological culture of the mid-third millennium BCE, and by implication with the later extensions of this culture. In **8.4** an attempt is made to test this against the archaeological evidence, particularly the archaeobotanical evidence, and the fit is found to be reasonable though not perfect. Reasoning from the current locations of Dravidian languages, as well as from the records of Dravidian-speaking kingdoms in South India in the early centuries of the CE, it would seem unsurprising that the largest complex known to archaeology in South India had the same linguistic affiliation. However, one may ask if this hypothesis can be *dis*proved: is there any archaeological evidence of discontinuity which might suggest that Dravidian speakers entered the region in the intervening period?

Chapter 8 also infers that Proto-Dravidian was spoken, either earlier or contemporaneously, in the Krishna–Godavari area. This would suggest a search for

possible links between the archaeological sites in the two regions. It has already been pointed out that there are shared elements between some Southern Neolithic sites and sites in the northern Deccan, which can be associated with Dravidian elements in the Marathi language (Chapter 9, note 1) as well as Dravidian place names in Maharashtra (Chapter 9). A further important inference of Chapter 8 is that the reconstructed language known as Proto-South Dravidian (PSD) represents an advanced culture with urban structures, complex irrigation works, rulers and administrators, armies and navies, taxation and customs duties, etc. (**8.37**). Even assuming that PSD, like all reconstructed languages, represents a composite of speech over a long period of time and over a wide geographical area, it is still striking that there is no archaeological site remotely resembling this description in South India in the period to which this reconstruction can be taken to apply, that is, the late second millennium BCE at the latest. Thus either our linguistic chronology is wildly off, or there are sites waiting to be discovered. Such sites might be only large towns – or even large villages, or groups of villages – from our present perspective, and might be located in areas which are as yet unexplored, particularly coastal areas. Or they might be linguistic phantoms.

Chapter 9 explores the place names of Maharashtra and finds a large number of them to be of Dravidian origin. Combined with evidence of Dravidian elements in the Marathi language, and archaeological connections between sites in Maharashtra and farther south, a strong case can be made for a previous Dravidian-speaking population in Maharashtra. This is supported by ethnographic and historical evidence. Similar place names are also found in Gujarat and (probably) Sindh, supporting other evidence for an early Dravidian presence in those areas, possibly during the time of the Indus Valley culture (see earlier text). What is perhaps most striking is the concentration of Dravidian place names in coastal Maharashtra, which considered along with later evidence for South Indian involvement in sea trade (**8.42**), leads to the possibility that sea trade was perhaps much earlier in peninsular India than it is now thought to be.

The examples given here are some of the kinds of inferences produced by linguistic archaeology, along with suggestions about their possible relevance to archaeology. Testing linguistic inferences against archaeological evidence can be useful in two ways. For one, it can provide a test of the methods and assumptions of linguistic archaeology, and lead to their refinement. For example, we know that the uniform appearance of a reconstructed language may conceal variants which in reality occurred over a long period of time and over a large spatial area, but we have no idea of what the temporal and spatial *limits* might be, if any. By confronting the reconstructed language with an archaeological entity which partly corresponds to it, we can begin to see what these limits might be: thus, the comparison of Proto-Dravidian with descriptions of the Southern Neolithic (**8.41**) notes a number of reconstructed words for items which do not turn up archaeologically, such as two-storied buildings, ladders/staircases, wheels, and axles. Erring in the other direction, the PD reconstructions fail to include words for

a number of millets found (sporadically) at the Southern Neolithic sites, though such words are found in the next stage, PSD.

On the other hand, when the linguistic results are found to fit with the archaeological findings, then the archaeologist and linguist are in a position to collaborate in developing a more complete picture of the prehistoric past. Admittedly, archaeological research can be done, and is done, in the absence of any notion of the subject population's linguistic repertoire. Yet language is the glue that holds speech communities together, and when linguistic information is available and can be made to fit with the archaeological evidence, the result can be a richer and more detailed picture of prehistory.

Notes

1 In ancient South Asia, though the Prakrits provide some suggestions of regional dialect variation (see **2.82**), we really have no knowledge of the relationship between the literary tradition that has been handed down to us, and the actual usage of the majority of Indo-Aryan speakers. The evidence of Chapters 5 and 6 suggests that regional variation was probably greater, even from the earliest times, than we can infer from any analysis of the traditional texts.

2 The Nuristani or Kafiri languages, a separate branch of the Indo-Iranian languages, may have found their way to their current locations a few centuries earlier.

3 That is to say, languages related to Dravidian which were later lost, and thus could not be included in the reconstructions of Proto-Dravidian.

4 Note that Korku, a North Munda language, is spoken in Nimar District of Madhya Pradesh (Kuiper 1962: 43).

5 If we look for a comparable situation, we might think of the Ottoman Empire: when Turkish was no longer the administrative language in the Balkans, the local languages remained, albeit with a certain amount of Turkish content.

6 Speakers of outer IA may also have entered the Kosala/Avadh area from the Narmada across the Vindhya complex, via the valleys of the Son and other rivers.

7 For goddess worship in Maharashtra see Feldhaus (1995, 1996) with bibliography; for Bengal see McDermott (2000), Sarkar 2001. For serpent worship see Vogel (1926), Panda (1986), Sāmant and Kabirāj (2002). See Misra (2001: 518) for goddess worship in the Jorwe culture of Maharashtra (1500–900 BCE).

BIBLIOGRAPHY

(including bibliographical abbreviations)

Algeo, J. (1972) *Problems in the Origins and Development of the English Language* (2nd edition), New York: Harcourt Brace Jovanovitch.

Allchin, B. and F. R. Allchin (1982) *The Rise of Civilization in India and Pakistan*, Cambridge: Cambridge University Press.

Allchin, F. R. (1963) *Neolithic Cattle-keepers of South India: A Study of the Deccan Ashmounds*, Cambridge: Cambridge University Press.

—— (1979a) "Early domesticated animals in India and Pakistan," in G. Possehl (ed.), *Ancient Cities of the Indus*, New Delhi: Vikas Publishing House, pp. 245–8.

—— (1979b) "Early cultivated plants in India and Pakistan," in G. Possehl (ed.), *Ancient Cities of the Indus*, New Delhi: Vikas Publishing House, pp. 249–52.

—— (1995) *The Archaeology of Early Historic South Asia: The Emergence of Cities and States*, Cambridge: Cambridge University Press.

Andronov, M. (1969) *A Standard Grammar of Modern and Classical Tamil*, Madras: New Century Book House.

Apte, N. G. (1967) *Mahārāṣṭra Rājya Grāmasūci* [Maharashtra state village-list], Pune: Mahārāṣṭra Grāmakoś Maṇḍaḷ [Maharashtra Village-Lexicon Society].

Ayyar – see Krishna Ayyar

Bagchi, P. C. (1975) *Pre-Aryan and Pre-Dravidian in India*, Calcutta: Calcutta University Press.

Banerjea, J. N. (1960) *J. N. Banerjea Volume: a collection of articles by his friends and pupils presented on his retirement from Carmichael professorship of ancient Indian history and culture, Univ. of Calcutta*, Calcutta: University of Calcutta.

Barth, F. (1969) "Pathan identity and its maintenance," in F. Barth (ed.), *Ethnic Groups and Boundaries*, Boston, MA: Little, Brown & Co., pp. 117–34.

Basham, A. L. (1954) *The Wonder that was India*, New York: Grove Press, Inc.

Baugh, A. C. (1957) *A History of the English Language* (2nd edition), New York: Appleton-Century-Crofts.

Baugh, A. C. and T. Cable (1993) *A History of the English Language* [4th edition], Englewood Cliffs, NJ: Prentice-Hall.

Bauman, R. and J. Sherzer (1974) *Explorations in the Ethnography of Speaking*, Cambridge: Cambridge University Press.

Bedigian, D. and J. H. Harlan (1986) "Evidence for the cultivation of sesame in the ancient world," *Economic Botany* **40**: 137–54.

Beekes, S. P. R. (1995) *Comparative Indo-European Linguistics: An Introduction*, Amsterdam: John Benjamins Publishing.

Berger, H. (1955) *Zwei Probleme der mittelindischen Lautlehre*, München: J. Kitzinger.

Bhagwat, M. M. (1967) The Varhadi dialect: a study of Varhadi dialect in the neighbourhood of Akola, unpublished PhD dissertation, Deccan College Postgraduate and Research Institute, Poona.

Bhandarkar, Sir R. G. (1927) *Collected Works of Sir R. G. Bhandarkar, Vol. 3: Early History of the Deccan*, ed. by N. B. Utgikar, Poona: Bhandarkar Oriental Research Institute.

Blake, N. F. (ed.) (1992) *The Cambridge History of the English Language*, Vol. II: 1066–1476, Cambridge, New York: Cambridge University Press.

Bloch, J. (1919) *La formation de la langue marathe*, Paris: Librairie Ancienne Honoré Champion.

—— (1925) "Sanskrit et Dravidien," *Bulletin de la Société Linguistique* **25**: 1–21. (English translation by P. Bagchi, in Bagchi, P. C. (1975) *Pre-Aryan and Pre-Dravidian in India*, Calcutta: Calcutta University Press, pp. 35–62.)

—— (1950) *Les inscriptions d'Asoka*, Paris: Société d'Édition Les Belles Lettres.

—— (1965) *Indo-Aryan from the Vedas to Modern Times*, Paris: Librarie d'Amérique et d'Orient Adrien-Maisonneuve–Jean Maisonneuve Succ.

Bloomfield, L. (1933) *Language*, New York: Holt and Company.

Blust, R. (1996a) "Beyond the Austronesian homeland: The Austric hypothesis and its implications for archaeology," in W. H. Goodenough (ed.), *Prehistoric Settlement of the Pacific*, Philadelphia, PA: *TAPS* **86**.5: pp. 117–40.

Bökönyi, S. (1997) "Horse remains from the prehistoric site of Surkotada, Kutch, late 3rd millennium B.C." *South Asian Studies* **13**: 297–307.

Bonvilain, N. (2000) *Language, Culture, and Communication: The Meaning of Messages*, Upper Saddle River, NJ: Prentice Hall.

Bray, D. de S. (1909) *The Brahui Language. Part 1, Introduction and Grammar*, Calcutta: Superintendent of Government Printing.

Bright, W. (1972) "The enunciative vowel," *IJDL* **1**: 26–55.

—— (1975) "The Dravidian enunciative vowel," in H. Schiffman and C. Eastman (eds), *Dravidian Phonological Systems*, Seattle: South Asian Studies Program, Institute for Comparative and Foreign Area Studies, University of Washington, pp. 11–46. (Reprinted in Bright (1990) *Language Variation in South Asia*, New York: Oxford University Press, pp. 86–117.)

—— (1990) *Language Variation in South Asia*, New York: Oxford University Press.

Bronkhorst, J. and M. M. Deshpande (eds) (1999) *Aryan and Non-Aryan in South Asia: Evidence, Interpretation and Ideology*, Cambridge, MA: Harvard University Press.

BSOAS = *Bulletin of the School of Oriental and African Studies*, University of London.

Buck, C. D. (1933) *Comparative Grammar of Greek and Latin*, Chicago, IL: University of Chicago Press.

Bulcke, C. (1981) Angrezī–Hindī koš [English–Hindi dictionary], New Delhi, Chanda.

Burrow, T. (1945) "Some Dravidian words in Sanskrit," *TPS*, pp. 79–100.

—— (1946) "Loan words in Sanskrit," *TPS*, pp. 1–30.

—— (1947a) "Dravidian Studies vii: further Dravidian words in Sanskrit," *BSOAS* **12**: 365–96.

—— (1947b) "Dravidian Studies vi: the loss of initial c/s in Dravidian," *BSOAS* **12**.

—— (1955) *The Sanskrit Language*, London: Faber & Faber.

—— (1971) "Spontaneous cerebrals in Sanskrit," *BSOAS* **34**: 538–59.

——(1972) "The primitive Dravidian word for the horse," *IJDL* **1**: 1.18–25.

—— (1973a) *The Sanskrit Language* (2nd edition), London: Faber & Faber.
—— (1973b) 'The Proto-Indo-Aryans', *JRAS* **1**: 123–40.
Burrow, T. and M. B. Emeneau (1968) *A Dravidian Etymological Dictionary: Supplement*, Oxford: Oxford University Press.
—— (1972) "Dravidian etymological notes," *JAOS* **92**: 397–418, 475–91.
—— (1984) *A Dravidian Etymological Dictionary* (2nd edition), Oxford: Clarendon Press.
Caillat, C. (ed.) (1989a) *Dialects dans les littératures indo-aryennes*, Paris: Institut de Civilisation Indienne.
—— (1989b) "Sur l'authenticité linguistique des édits d'Asoka," in C. Caillat (ed.), *Dialects dans les litteratures indo-aryennes*, Paris: Institut de Civilisation Indienne, pp. 413–32.
Callary, R. E. (1971) "Dialectology and linguistic theory," *American Speech* **46**: 200–09.
Cardona, G. (1965) *A Gujarati Reference Grammar*, Philadelphia, PA: University of Pennsylvania Press.
—— (1974) "The Indo-Aryan languages," *Encyclopedia Britannica* (15th edition), pp. 439–50.
Casal, J.-M. (1961) *Fouilles de Mundigak*, Paris: Librairie C. Klinksieck.
CDIAL = Turner, Sir R. L. (1966) *A Comparative Dictionary of the Indo-Aryan Languages*, London: Oxford University Press.
Chambers, J. K. and P. Trudgill (1998) *Dialectology* (2nd edition), Cambridge: Cambridge University Press.
Chatterjee, P. G. (1985) "The role and impact of nature on culture: Resplendent results from the Indian neolithic," in M. Bellwood (ed.), *Recent Advances in Indo-Pacific Prehistory: Proceedings of the International Symposium held at Poona, December 19–21, 1978*, New Delhi: Oxford & IBH Publishing Co.
Chatterji, S. K. (1970) *The Origin and Development of the Bengali Language*, London: George Allen & Unwin Ltd. (First published by Calcutta University Press, 1926.)
Chitnis, V. S. (1964) The Khandeshi dialect, as spoken by the farmers of Mohadi in the Dhulia Taluka, unpublished PhD dissertation, University of Poona.
CIIL = Central Institute of Indian Languages, Manasagangotri, Mysore, India.
Collinder, B. (1977) *Fenno-Ugric Vocabulary; An Etymological Dictionary of the Uralic Languages* (2nd edition), Stockholm: Almqvist & Wiksell.
Cotéaunu, J. (1957) "À propos des langues mixtes (sur l'istro-roumain)," in *Mélanges linguistiques publiés à l'occasion du VIIIe. Congrès International des Linguistes à Oslo, du 5 au 9 Août 1957*, Bucarest: Éditions de l'Académie Populaire Roumaine, pp. 129–48.
Dalai, U. P. (1998) "The Dravidian substratum effect on the Oriya boli-type construction," *IJDL* **27**: 85–96.
Dalton-Puffer, C. (1996) *The French Influence on English Morphology: A Corpus-based Study of Derivation*, Berlin, New York: Mouton de Gruyter.
DBIA = Emeneau, M. B. and T. Burrow (1962) *Dravidian Borrowings from Indo-Aryan*, Berkeley, CA: University of California Press.
DEDR = Burrow, T. and M. B. Emeneau (1984) *A Dravidian Etymological Dictionary* (2nd edition), Oxford: Clarendon Press.
DEDS = Burrow, T. and M. B. Emeneau (1968) *A Dravidian Etymological Dictionary: Supplement*, Oxford: Oxford University Press.
Degener, A. (2002) "The Nuristani languages," in N. Sims-Williams (ed.), *Indo-Iranian Languages and Peoples*, London: The British Academy.
DEN = T. Burrow and M. B. Emeneau (1972) "Dravidian etymological notes," *JAOS* **92**: 397–418, 475–91.

Deo, S. B. (1974) *Excavations at Bhokardan (Bhogavardhana), 1973*, Nagpur: Nagpur University.
—— (1982) *Recent Researches on the Chalcolithic and Megalithic Cultures of the Deccan*, Madras: University of Madras.
Deshpande, C. D. (1971) *Geography of Maharashtra*, New Delhi: National Book Trust.
Deshpande, M. M. (1978) *Paninian Grammarians on Dialectal Variation in Sanskrit*, Madras: Adyar Library Bulletin.
—— (1979a) "Genesis of Rigvedic retroflection: A historical and sociolinguistic investigation," in M. Deshpande and P. E. Hook (eds), *Aryan and non-Aryan in India*, Ann Arbor, MI: The University of Michigan Center for South and Southeast Asian Studies, pp. 235–306.
—— (1979b) *Sociolinguistic Attitudes in India: An Historical Reconstruction*, Ann Arbor, MI: Karoma Publishers.
Deshpande, M. M. and P. E. Hook (eds) (1979) *Aryan and non-Aryan in India*, Ann Arbor: The University of Michigan Center for South and Southeast Asian Studies.
De Silva, M. W. S. (1965) A structural analysis of the Vedda language spoken in the Badulla and Polonnaruwa Districts of Ceylon, unpublished PhD thesis, University of London.
—— (1972) *Vedda Language of Ceylon: Texts and Lexicon*, München: R. Kitzinger.
Devaraj, D. V., J. G. Shaffer, C. S. Patil and Balasubramanya (1995) "The Watgal excavations: an interim report," *Man and Environment* **20** (2): 57–74.
Dil, A. S. (ed.) (1980) *Language and Linguistic Area: Essays by Murray B. Emeneau*, Stanford, CA: Stanford University Press.
Dines, E. R. (1980) "Variation in discourse – 'and stuff like that' ," *Language in Society* **9**: 13–31.
EJVS = *Electronic Journal of Vedic Studies* <http://users.primushost.com/~india/ejvs/>
Ekka, F. (1972) "Men's and women's speech in Kūṛux," *Linguistics* **81**: 25–31.
Emeneau, M. B. (1943) "A Dravidian etymology of the Sanskrit proper name Nala," in *University of California Publications in Classical Philology*, pp. 255–62.
—— (1954) "Linguistic prehistory of India," *PAPS* **98**: 282–92.
—— (1956) "India as a linguistic area," *Language* **32**: 3–16.
—— (1961) "North Dravidian velar stops," in M. A. Palamurukan (ed.), *Studies Presented to Professor T. P. Meenakshisundaram*, Coimbatore, pp. 371–88.
—— (1962a) *Brahui and Dravidian Comparative Grammar*, Berkeley and Los Angeles, CA: University of California Press.
—— (1962b) "Bilingualism and structural borrowing," *PAPS **106**:* 430–42. (Reprinted in Dil, A. S. (ed.) (1980) *Language and Linguistic area: Essays by Murray B. Emeneau*, Stanford, CA: Stanford University Press, 38–65.)
—— (1966) "The dialects of Old Indo-Aryan," in H. Birnbaum and J. Puhvel (eds), *Ancient Indo-European Dialects*, Berkeley and Los Angeles, CA: University of California, pp. 123–38.
—— (1967) 'The South Dravidian languages', *JAOS* **87**: 365–413.
—— (1970) *Dravidian Comparative Phonology: A Sketch*, Annamalainagar: Annamalai University.
—— (1971) "Dravidian and Indo-Aryan: The Indian linguistic area," in A. F. Sjoberg (ed.), *Symposium on Dravidian civilization*, Austin, TX and New York: Jenkins Publishing Company, The Pemberton Press. (Reprinted in Dil, A. S. (ed.) (1980) *Language and Linguistic Area: Essays by Murray B. Emeneau* Stanford, CA: Stanford University Press, pp. 167–96.)

Emeneau, M. B. and T. Burrow (1962) *Dravidian Borrowings from Indo-Aryan*, Berkeley, CA: University of California Press.

Erdosy, G. (ed.) (1995a) *The Indo-Aryans of Ancient South Asia: Language, Material Culture, and Ethnicity*, Berlin and New York: Walter de Gruyter.

—— (1995b) "Language, material culture, and ethnicity," in G. Erdosy (ed.), *The Indo-Aryans of Ancient South Asia: Language, Material Culture, and Ethnicity*, Berlin and New York: Walter de Gruyter, pp. 1–31.

Fairservis, W. A. (1992) *The Harappan Civilization and its Writing: A Model for the Decipherment of the Indus Script*, Delhi: Oxford & IBH.

Falk, H. (1986) *Bruderschaft und Würfelspiel: Untersuchungen zur Entwicklungsgeschichte des vedischen Opfers*, Freiburg: H. Falk.

Farmer, Steve, Richard Sproat and Michael Witzel (forthcoming in *EJVS*) "The collapse of the Indus-script thesis: the myth of a literate Harappan civilization" <http://users.primushost.com/~india/ejvs>

Feldhaus, A. (1983) *The Religious System of the Mahānubhāva Sect: The Mahānubhāva Sūtrapāṭha*, New Delhi: Manohar.

—— (1995) *Water and Womanhood: Religious Meanings of Rivers in Maharashtra*, New York: Oxford University Press.

—— (ed.) (1996) *Images of Women in Maharashtrian Literature and Religion*, Albany: State University of New York Press.

Ferguson, C. (1964) "Diglossia," in D. Hymes (ed.), *Language in Culture and Society: A Reader in Linguistics and Anthropology*, New York: Harper & Row Publishers, pp. 429–37.

Fischer, O. (1992) "Syntax," in N. F. Blake (ed.), *The Cambridge History of the English Language*, Vol. II: 1066–1476, Cambridge and New York: Cambridge University Press, pp. 207–408.

Fox, A. (1995) *Linguistic Reconstruction: An Introduction to Theory and Method*, Oxford and New York: Oxford University Press.

Friedrich, P. (1970) *Indo-European Trees: The Arboreal System of a Prehistoric People*, Chicago, IL: University of Chicago Press.

Fuller, D. Q. (2001a) "Brahmagiri and beyond: the archaeology of the Southern Neolithic," in S. Settar and R. Korisettar (eds), *Indian Archaeology in Retrospect*, Vol. I, *Prehistory: Archaeology of South Asia*, Delhi: Manohar, in association with Indian Council of Historical Research.

—— (2001b) "An agricultural perspective on Dravidian historical linguistics: archaeological crop packages, livestock and Dravidian crop vocabulary," in C. Renfrew and P. Bellwood (eds), *Examining the Farming/language Dispersal Hypothesis*, Cambridge: McDonald Institute for Archaeological Research.

Fuller, D., R. Korisettar and P. C. Venkatasubbiah (2001) 'Southern neolithic cultivation systems: a reconstruction based on archaeobotanical evidence', *South Asian Studies* **17**: 171–87.

Gardner, P. (1980) "Lexicostatistics and Dravidian differentation *in situ*," *Indian Linguistics* **41**: 170–80.

Garren, W. R. and C. R. Page (1983) *Gazetteer of Pakistan: Names Approved by the United States Board of Geographic Names*, Washington, DC: Defense Mapping Agency.

Geiger, W. (1938) *A Grammar of the Sinhalese Language*, Colombo: The Royal Asiatic Society, Ceylon Branch.

Genetti, C. (1988) "A contrastive study of the Dolakhali and Kathmandu Newari dialects," *Cahiers linguistiques de l'Asie centrale* **17**: 161–91.

George, K. M. (1986) *Place Names of Southern India*, Trivandrum: Dravidian Linguistic Association.

Ghatage, A. M. (1964) *Phonemic and Morphemic Frequencies in Hindi*, Poona: Deccan College Postgraduate and Research Institute.

—— (1966) *A Survey of Marathi Dialects III: Kuṇabī of Mahāḍ*, Bombay: State Board for Literature and Culture.

—— (1967) *A Survey of Marathi Dialects IV: Cochin*, Bombay: State Board for Literature and Culture.

—— (1968a) *A Survey of Marathi Dialects V: Konkani of Kankon*, Bombay: State Board for Literature and Culture.

—— (1968b) *A Survey of Marathi Dialects VII: Wārlī of Thana*, Bombay: State Board for Literature and Culture.

—— (1970) *A Survey of Marathi Dialects VI: Marathi of Kasargod*, Bombay: State Board for Literature and Culture.

—— (1972) *A Survey of Marathi Dialects VIII: Gāwḍi of Goa*, Bombay: State Board for Literature and Culture.

Goswami, G. C. (1966) *An Introduction to Assamese Phonology*, Pune: Deccan College Postgraduate and Research Institute.

Gramasuci = Apte, N. G. (1967) *Mahārāṣṭra Rājya Grāmasūci* [Maharashtra state village- list], Pune: Mahārāṣṭra Grāmakoś Maṇḍaḷ [Maharashtra Village-Lexicon Society].

Grierson, Sir G. A. (1931–33) "On the modern Indo-Aryan vernaculars," *The Indian Antiquary* **60–62**.

—— (1967) (LSI) *Linguistic Survey of India*, Delhi: Motilal Banarsidass. (Originally published in Calcutta, 1903–28.)

Gudschinsky, S. (1956) "The ABC's of lexicostatistics (glottochronology)," *Word* **12**: 175–210.

Gumperz, J. J. (1958) "Dialect differences and social stratification in a north Indian village," *American Anthropologist* **60**: 668–82.

—— (1982a) *Language and Social Identity*, Cambridge, Melbourne, etc.: Cambridge University Press.

—— (1982b) *Discourse Strategies*, Cambridge, Melbourne, etc.: Cambridge University Press.

Gumperz, J. J. and D. Hymes (1972) *Directions in Sociolinguistics*, New York, Chicago, IL, etc.: Holt, Rinehart and Winston, Inc.

Gumperz, J. J. and R. Wilson (1971) "Creolization without pidginization: a case from the Indo-Aryan/Dravidian border," in D. Hymes (ed.), *Pidginization and Creolization of Languages*, Cambridge: Cambridge University Press, pp. 151–67.

Haas, M. (1944) "Men's and women's speech in Koasati," *Language* **20**: 142–9.

Haeri, N. (1996) *The Sociolinguistic Market of Cairo: Gender, Class and Education*, London: Kegan Paul International.

Hall, R. A. Jr. (1976) *Proto-Romance Phonology*, New York: Elsevier.

Hallock, R. (1962) "The pronominal suffixes in Achaemenid Elamite," *Journal of Near Eastern Studies* **23**: 53–6.

Hardgrave, R. (1965) *The Dravidian Movement*, Bombay: Popular Prakashan.

Harlan, J. R. (1971) "Agricultural origins: centers and noncenters," *Science* **174**: 468–74.

Haugen, E. (1950) "The analysis of linguistic borrowing," *Language* **26**: 210–31.

Heath, J. (1978) *Linguistic Diffusion in Arnhem Land*, Canberra: Australian Institute of Aboriginal Studies.

—— (1981) "A case of intensive lexical diffusion," *Language* **57:** 335–67.

Hock, H. H. (1975) "Substratum influence on (Rig Vedic) Sanskrit?," *Studies in the Linguistic Sciences* **5**: 76–125.

—— (1979) "Retroflection rules in Sanskrit," *South Asian Language Analysis* **1**: 47–62.

—— (1993) "Subversion or convergence? The issue of Pre-Vedic retroflection re-examined," *Studies in the Linguistic Sciences* **23**: 73–115 [publ. 1996].

Hock, H. H. and R. Pandharipande (1976) "The sociolinguistic position of Sanskrit in pre-Muslim South Asia," *Studies in language learning* **1**: 102–38.

Hockings, P. (1980) *Ancient Hindu Refugees: Badaga Social History, 1550–1975*, New Delhi: Vikas Publications.

Hogg, R. M. (ed.) (1992) *The Cambridge History of the English Language*, Vols. I–VI, Cambridge and New York: Cambridge University Press.

Houston, A. (1985) "Establishing the continuity between past and present morphology," Chapter 6 in Continuity and change in English morphology: the variable (ING), University of Pennsylvania PhD dissertation, 220–86.

—— (1991) "A grammatical continuum for (ING)," in P. Trudgill and J. Chambers (eds), *Dialects of English: Studies in Grammatical Variation*, Singapore: Longman Singapore, pp. 241–57.

Hymes, D. (ed.) (1964) *Language in Culture and Society: A Reader in Linguistics and Anthropology*, New York: Harper & Row Publishers.

—— (ed.) (1971) *Pidginization and Creolization of Languages* Cambridge: Cambridge University Press.

IIJ = *Indo-Iranian Journal*.

IJAL = *International Journal of American Linguistics*.

IJDL = *International Journal of Dravidian Linguistics*.

India, Government of (1999) *Census of India 1991: Maharashtra State District Profile 1991*, Delhi: Controller of Publications.

Irschick, E. (1969) *Politics and Social Conflict in South India: The Non-Brahman Movement and Tamil Separation, 1916–1929*, Berkeley, CA: University of California Press.

Jankowsky, K. R. (1990) "The neogrammarian hypothesis," in E. C. Polomé (ed.), *Research Guide on Language Change*, Berlin: Mouton de Gruyter, pp. 223–39.

JAOS = *Journal of the American Oriental Society*.

Jarrige, J. F. (1979) "Excavations at Mehrgarh, Pakistan," in J. E. Van Lohuizen-de Leeuw (ed.), *South Asian Archaeology 1975*, Leiden: E. J. Brill, pp. 76–87.

—— (1982) "Excavations at Mehrgarh: their significance for understanding the background of the Harappan civilization," in G. Possehl (ed.), *Harappan Civilization*, New Delhi: Oxford and IBH Publishing Co., in collaboration with the American Institute of Indian Studies.

Jazayery, M. A. and W. Winter (eds) (1988) *Languages and Cultures: Studies in Honor of Edgar C. Polomé*, Berlin: Mouton De Gruyter.

Jha, S. (1958) *The Formation of the Maithili Language*, London: Luzac & Company.

Joshi, S. B. (1951) "Etymology of place-names paṭṭi-haṭṭi, some observations on the history of Maharashtra and Karnataka," *Annals of the Bhandarkar Oriental Research Institute* **32**: 41–56.

JRAS = *Journal of the Royal Asiatic Society*.

Kalogjera, D. (1991–92)" 'Rudimentary' bilingualism (a report on work in progress)," *Studia Romanica et Anglica Zagrebiensia* 36–37, 63–75.

Katre, S. M. (1966) *The Formation of Konkani*, Poona: Deccan College Post-graduate and Research Institute.

Katre, S. M. (1968) *Problems of Reconstruction in Indo-Aryan*, Simla: Indian Institute of Advanced Study.

Kavadi, N. B. and F. C. Southworth (1965) *Spoken Marathi I*, Philadelphia, PA: University of Pennsylvania Press.

Kelkar, A. R. (1957) " 'Marathi English': a study in foreign accent," *Word* **13**: 268–92.

—— (1958) *The Phonology and Morphology of Marathi*, Ann Arbor: University Microfilms.

Kennedy, K. A. R. (1995) "Have Aryans been identified in the prehistoric skeletal record from South Asia? Biological anthropology and concepts of ancient races," in G. Erdosy (ed.), *The Indo-Aryans of Ancient South Asia: Language, Material Culture and Ethnicity*, Berlin and New York: Walter de Gruyter, pp. 32–66.

Kenoyer, J. M. (1989) *Old Problems and New Perspectives in the Archaeology of South Asia*, Madison, WI: Department of Anthropology, University of Madison.

Klaiman, M. H. (1977) "Bengali syntax: possible Dravidian influence," *IJDL* **6**: 303–17.

Kocchar, S. L. (1998) *Economic Botany in the Tropics*, Delhi: MacMillan India Limited.

Kosambi, D. D. (1956) *An Introduction to the Study of Indian History*, Bombay: Popular Book Depot.

Krishna Ayyar, K. V. (1966) *A Short History of Kerala*, Ernakulam: Pai Publications.

Krishnamurti, B. (1969) "Comparative Dravidian studies," in T. Sebeok (ed.) (1969) *Current Trends in Linguistics*, Vol. 5: *Linguistics in South Asia*, Paris: Mouton, pp. 309–33.

—— (1975) "Gender and number in Proto-Dravidian," *IJDL* **4**: 328–50.

—— (1978) "Areal and lexical diffusion of sound change: evidence from Dravidian," *Language* **54**: 1–20.

—— (1980) "An overview of comparative Dravidian studies since Current Trends 5," Presidential lecture delivered to the 10th Annual Conference of the Dravidian Linguistic Association, New Delhi.

—— (2002) "Linguistic prehistory and borrowing," in V. Swarajyalakshmi (ed.), *Case for Language Studies: Papers in Honour of Professor B. Laksmi Bai*, Hyderabad: Centre of Advanced Study in Linguistics, Osmania University/Book Links Cooperation, pp. 199–208.

—— (2003) *The Dravidian Languages: A Comparative, Historical and Typological Study*, Cambridge: Cambridge University Press.

Kuiper, F. B. (1948) "Proto-Munda words in Sanskrit." *Verhandelingen der Kon. Nederl. Akademie van Wetenschappen, Afd. Letterkunds* **51:3**.

—— (1955) "Rigvedic loanwords," in O. Spies (ed.), *Studia Indologica: Festschrift für Willibald Kirfel*, Bonn: Orientalistisches Seminar der Universität Bonn, pp. 137–85.

—— (1962) *Nahali: A Comparative Study*, Amsterdam: N. V. Noord-Hollandsche Uitgevers Maatschappij.

—— (1967) "The genesis of a linguistic area," *Indo-Iranian Journal* **10**: 81–102.

—— (1991) *Aryans in the Rigveda*, Amsterdam and Atlanta, GA: Rodopi.

Kulkarni, S. B. (1969) Descriptive analysis of Katkari dialect, unpublished PhD thesis in linguistics, University of Poona, Poona.

—— (1976) *Bhili of Dangs*, Poona: Deccan College.

—— (n.d.) "Kāṭkari lok va tyāncyā bhāśetil paśupakṣivācak śabd" ['The Katkari people and their words for animals'], *Journal of the University of Poona* **23**: 91–121.

Labov, W. (1963) "The social motivation of a sound change," *Word* **19**: 273–309.

—— (1966) *The Social Stratification of English in New York City*, Washington, DC: Center for Applied Linguistics.

—— (1972) *Sociolinguistic Patterns*, Philadelphia, PA: University of Pennsylvania Press.

—— (1980) *Locating Language in Time and Space: Quantitative Analyses of Linguistic Structure*, New York, London, etc.: Academic Press.

—— (1981) 'Resolving the neogrammarian controversy', *Language* **57**: 267–308.

—— (2001) *Principles of Linguistic Change, Vol. 2: Social Factors,* Oxford & Cambridge: Blackwell Publishers Inc.

Lal, B. B. (1954) *Kashmir through the Ages, 5000* B.C. *to 1954* A.D*; A Historical Survey*, Srinagar, Kashmir: Chronical Pub. House.

—— (1971) 'Perhaps the earliest ploughed field so far excavated anywhere in the world', *Puratattva* **4**: 20–32.

Lamberg-Karlovsky, C. (2003) "Archaeology and language: the Indo-Iranians," *Current Anthropology* **43**: 63–88.

Lee, G. (1986) "Diglossia in ancient India," *Ohio State University Working Papers in Linguistics* **34**: 151–64.

Lévi, S. (1923) "Pré-aryen et pré-dravidien dans l'Inde," *Journal Asiatique* **203**: 1–57. (English translation by P. C. Bagchi in P. C. Bagchi, (1975) *Pre-Aryan and Pre-Dravidian in India*, Calcutta: Calcutta University Press.)

Lorimer, D. L. R. (1935–38) *The Burushaski Language*, Vols. 1–3, Oslo: H. Aschenhoug.

McAlpin, D. W. (1981) *Proto-Elamo-Dravidian: The Evidence and its Implications*, Philadelphia, PA: The American Philosophical Society.

McCormack, W. C. and S. A. Wurm (eds) (1979) *Language and Society: Anthropological Issues* The Hague–Paris–New York, Mouton Publishers.

McCrindle, J. W. (1877) *Ancient India as described by Megasthenes and Arrian*, Calcutta: Thacker, Spink; London: Trübner.

McDavid, R. (1948) "Postvocalic /r/ in South Carolina: a social analysis," *American Speech* **23**: 194–203.

McDermott, R. F. (2000) *Mother of my Heart, Daughter of my Dreams: Transformations of Kali and Uma in the Devotional Poetry of Bengal*, New York: Oxford University Press.

MacDonell, A. A. (1927) *A Sanskrit Grammar for Students*, London: Oxford University Press.

Mahadevan, I. (1981–82) 'S. R. Rao's decipherment of the Indus script', *Puratattva* **9**: 34–42.

Majumdar, P. C. (1970) *A Historical Phonology of Oriya*, Calcutta: Sanskrit College.

Mallory, J. P. (1989) *In search of the Indo-Europeans: Language, archaeology and myth.* London: Thames and Hudson.

Marshall, Sir John (1931) *Mohenjo-Daro and the Indus Civilization*, Vols. 1–3. London: Arthur Probstain.

Masica, C. P. (1976) *Defining a Linguistic Area: South Asia*, Chicago, IL: University of Chicago Press.

—— (1979) "Aryan and non-Aryan elements in north Indian agriculture," in M. M. Deshpande and P. E. Hook (eds), *Aryan and non-Aryan in India*, Ann Arbor: The University of Michigan Center for South and Southeast Asian Studies, pp. 55–152.

Masica, C. P. (1991) *The Indo-Aryan Languages*, Cambridge: Cambridge University Press.

Master, A. (1964) *A Grammar of Old Marathi*, Oxford: Clarendon Press.

Mayrhofer, M. (1951) "Arische Landnahme und indische Bevölkerung im Spiegel der altindischen Sprache," *Saeculum* **2**: 54–64.

—— (1953a) "Die Substrattheorien und das Indische," *Germanisch-Romanische Monatsschrift* **34**: 230–42.

—— (1953b) (M53) *Kurzgefasstes etymologisches Wörterbuch des Altindischen: A concise etymological Sanskrit dictionary*, Heidelberg: Carl Winter, Universitätsverlag.

—— (1974) *Die Arier im vorderen Orient – ein Mythos?* Mit einem bibliographischen Supplement. Wien: Verlag der Österreichischen Akademie der Wissenschaften, 1974

—— (1986) (M86) *Etymologisches Wörterbuch des Altindoarischen.*, Heidelberg: Carl Winter, Universitätsverlag.

Meadow, R. H. (1986) "Faunal exploitation in the greater Indus Valley: A review of recent work to 1980," in J. Jacobson (ed.), *Studies in the Archaeology of India and Pakistan*, New Delhi: Oxford and IBH, pp. 43–64.

—— (1987) "Faunal exploitation patterns in eastern Iran and Baluchistan: a review of recent investigations," in G. Gnoli and L. Lanciotti (eds), *Orientalia Iosephi Tucci Memoriae Dicata*, Rome: Istituto Italiano per il Medio ed Estremo Oriente, pp. 881–916.

—— (1989) "Continuity and change in the agriculture of the greater Indus Valley: the palaeobotanical and zooarchaeological evidence," in J. M. Kenoyer (ed.), *Old problems and new perspectives in the archaeology of South Asia*, Madison, WI: Department of Anthropology, University of Wisconsin, pp. 61–74.

Meadow, Richard and A. Patel (1997) "A comment on 'Horse remains from the prehistoric site of Surkotada, Kutch, late 3rd millennium B.C.' by Sandor Bökönyi," *South Asian Studies* **13**: 308–15.

Mehendale, A. M. (1948) *Historical Grammar of Inscriptional Prakrits*, Pune: Deccan College.

Mencher, J. P. (1974) "The caste system upside down, or the not so mysterious East," *Current Anthropology* **15**: 469–94.

—— (1976) "Politics, caste and religion in Madras villages: an analysis of their interactions," in G. R. Gupta (ed.), *Religion in Modern India*, New Delhi: Vikas Publications.

—— (1978) *Agriculture and Social Structure in Tamil Nadu: Past Origins, Present Transformations, and Future Prospects*, Bombay: Allied.

Menon, S. (1967) *A Survey of Kerala History*, Kottayam: National Book Stall.

Misra, H. (1975) *Historical Oriya Morphology*, Varanasi: Bharata Manisha Prakashan.

Misra, V. N. (2001) "Prehistoric human colonization of India," *Journal of Biosciences* **26** (suppl.): 491–531. Online. Available HTTP: <http://www.ias.ac.in/jbiosci/nov2001/533.pdf> (15 September 2003).

Misra, V. N. and P. Bellwood (1985) *Recent Advances in Indo-Pacific Prehistory: Proceedings of the International Symposium held at Poona, December 19–21, 1978*, New Delhi: Oxford and IBH Publishing Co.

Moag, R. (1994) *Malayalam – A University Course and Reference Grammar*, Austin, TX: Center for Asian Studies, University of Texas at Austin.

Molesworth, J. T. (1975) *Molesworth's Marathi–English dictionary*, Poona: Shubhada-Saraswat.

Monier-Williams, S. M. (1899) *A Sanskrit–English Dictionary*, Oxford, at the Clarendon Press.

Morgenstierne, G. (1973a) *Irano-Dardica*, Wiesbaden: L. Reichert: In Kommission bei O. Harrassowitz.

—— (1973b) *Indo-Iranian Frontier Languages*, Oslo: Universitetsforlaget.

Moylan, T. (1981) "Variation, change, and lexical diffusion in Oksapmin," paper read at the Annual Meeting of the Linguistic Society of America, December 1981.

—— (1982) System and subsystem, growth and decline, male and female: three pairs of concepts reflected in the process of language change in Oksapmin, Papua New Guinea, paper read at the Annual Meeting of the American Anthropological Association, November 1982.

Mulay, S. (1972) *Studies in the Historical and Cultural Geography and Ethnography of the Deccan, Based Entirely on the Inscriptions of the Deccan from the 1st–13th Century A.D.*, Poona: Deccan College Postgraduate and Research Institute.

Nadkarni, M. V. (1975) "Bilingualism and syntactic change in Konkani," *Language* **51**: 672–83.

Newman, Stanley (1956) "Vocabulary levels: Zuñi sacred and slang usage," *SWJA* **11**: 345–54 (reprinted in D. Hymes, ed. (1964), *Language in Culture and Society: A Reader in Linguistics and Anthropology*, New York: Harper & Row Publishers, pp. 397–406).

Niedermann, M. (1953) *Précis de phonétique historique du latin*, Paris: Librairie C. Klincksieck.

Nilakanta Sastri, K. A. (1955) *The Cōlas*, Madras: University of Madras.

—— (1964) *The Culture and History of the Tamils*, Calcutta: Firma K. L. Mukhopadhyay.

—— (1976) *A History of South India*, Madras: Oxford University Press.

Nurse, D. (1983) "A linguistic reconsideration of Swahili origins," *Azania* **18**: 127–50.

Nurse, D. and T. Spear (1985) *The Swahili: Reconstructing the History and Language of an African Society, 800–1500*, Philadelphia, PA: University of Pennsylvania Press.

Oberlies, T. (1994) *Review Article: F.B. J. Kuiper: Aryans in the Rigveda. IIJ* **37**: 333–349.

ODBL = Chatterji, S. K. (1970) *The Origin and Development of the Bengali Language*, London: George Allen & Unwin Ltd.

Paddayya, K. (1973) *Investigations into the Neolithic Culture of the Shorapur Doab, South India*, Leiden: Brill.

Panda, S. C. (1986) *Nāga Cult in Orissa*, Delhi: B. R. Publishing Corp.

Pandit, P. B. (1968) *India as a Sociolinguistic Area*, Gune Memorial Lectures, Pune: University of Pune.

—— (1972) *India as a Sociolinguistic Area*, Pune: University of Pune.

—— (1974) "Bilingual's grammar: a case study of Tamil–Saurashtri number names," in F. C. Southworth and M. L. Apte (eds), "Contact and convergence in South Asian languages," *IJDL* (special issue), pp. 181–97.

—— (1979) "Perspectives on sociolinguistics in India," in W. C. McCormack and S. A. Wurm (eds), *Language and Society: Anthropological Issues*, Mouton Publishers, the Hague–Paris–New York, pp. 171–82.

PAPS = *Proceedings of the American Philosophical Society.*

Parpola, A. (1988) "The coming of the Aryans to Iran and India and the cultural and ethnic identity of the Dāsas," *Studia Orientalia* **64**: 195–302.

—— (1994) *Deciphering the Indus Script*, Cambridge University Press.

—— (2002a) "From the dialects of Old Indo-Aryan to Proto-Indo-Aryan and Proto-Iranian," in N. Sims-Williams (ed.), *Indo-Iranian Languages and Peoples*, London: The British Academy.

—— (2002b) "Pre-Proto-Iranians of Afghanistan as initiators of Śākta Tantrism: on the Scythian/Saka affiliation of the Dāsas, Nuristanis and Magadhans," *Iranica Antiqua* **37**: 233–324.

Pattanayak, D. P. and G. N. Das (1972) *Conversational Oriya*, Mysore: Central Institute of Indian Languages.

Pedersen, H. (1931) *Linguistic Science in the Nineteenth Century*, Cambridge, MA: Harvard University Press. (Reissued as *The Discovery of Language*, Indiana University Press, 1959.)

Pinnow, H. J. (1953–54) 'Zu den altindischen Gewässernamen', *Beiträge zur Namensforschung* **4**: 217–34, **5**: 1–19.

—— (1959) *Versuch einer historischen Lautlehre der Kharia-Sprache*, Wiesbaden: Otto Harassowitz.

Pischel, R. (1965) *Life and Doctrines of the Buddha*, Vallabh Vidyanagar: Kapadia.

—— (1981) *Comparative Grammar of the Prakrit Languages*, Delhi: Motilal Banarsidass.

Pokorny, J. (1959) *Indogermanisches etymologisches Wörterbuch*, Bern: Francke.

Polomé, E. C. (ed.) (1990) *Research Guide on Language Change*, Berlin: Mouton de Gruyter.

Possehl, G. (1976) "Lothal: a gateway settlement of the Harappan civilization," in K. A. R. Kennedy and G. Possehl (eds), *Ecological Backgrounds of South Asian Prehistory*, Ithaca, NY: South Asia Program, Cornell University, pp. 118–31.

—— (1979) *Ancient Cities of the Indus*, New Delhi: Vikas.

—— (ed.) (1982) *The Harappan Civilization: A Contemporary Perspective*, New Delhi: Oxford & IBH Publishing Co.

—— (1996) *Indus Age: The Writing System*, New Delhi & Calcutta: Oxford & IBH Publishing Co. Pvt. Ltd.

—— (1999) *Indus Age: The Beginnings*, Philadelphia, PA: University of Pennsylvania Press.

Possehl, G. and K. A. R. Kennedy (1979) "Hunter-gatherer/agriculturalist exchange in prehistory: an Indian example," *Current Anthropology* **20**: 592–3.

Prins, A. A. (1952) *French Influence in English Phrasing*, Leiden: Press Universitaire.

Przyluski, J. (1926) "Un ancien peuple du Penjab: Les Udumbara," *Journal Asiatique* **206**: 25–36.

Pukui, M. K. and S. H. Elbert (1986) *Hawaiian Dictionary* [revised edition], Honolulu, HI: University of Hawaii Press.

Pulgram, E. (1958) *The Tongues of Italy, Prehistory and History*, Cambridge, MA: Harvard University Press.

Ramasubramaniam, K. S. (1968) *A Linguistic Study of Saurashtri Language*, Pune: University of Pune.

Randhawa, M. S. (1980) *A History of Agriculture in India*, New Delhi: Indian Council of Agricultural Research.

Randle, H. N. (1943) "An Indo-Aryan language of South India: Saurashtra Bhasa," *BSOAS* **XI**: 104–21.

Ratnagar, S. (1981) *Encounters, the Westerly Trade of the Harappa Civilization*, Delhi: Oxford University Press.

Rau, W. (1956) *Stadt und Gesellschaft im alten Indien*, Wiesbaden: Otto Harassowitz.

—— (1971) *Weben und Flechten im vedischen Indien.*, Wiesbaden: Fritz Steiner Verlag.

—— (1972) *Töpferei und Tongeschirr im vedischen Indien*, Wiesbaden: Fritz Steiner Verlag.

—— (1974) *Metale und Metalgeräte im vedischen Indien*, Wiesbaden: Fritz Steiner Verlag.

—— (1976) *The meaning of pur in Vedic literature*, München: Wilhelm Fink Verlag.

—— (1983) *Zur vedischen Altertumskunde*, Wiesbaden: Fritz Steiner Verlag.

—— (1997) "The earliest literary evidence for permanent Vedic settlements," in M. Witzel (ed.), *Inside the Texts, Beyond the Texts: New Approaches to the Study of the*

Vedas: Proceedings of the International Vedic Workshop, Harvard University, June 1989, Cambridge, MA: Harvard University Dept. of Sanskrit and Indian Studies; Columbia, MO : distributed by South Asia Books, pp. 203–6.

Renfrew, C. (1987) *Archaeology and Language: The Puzzle of Indo-European Origins*, London: Jonathan Cape (reviewed in Current Anthropology 29.3.437–68).

Rocher, L. (1986) *The Puranas*, Wiesbaden: Otto Harassowitz.

Saksena, B. (1971) *Evolution of Awadhi*, Delhi: Motilal Banarsidass.

Salah, A. S., V. M. Chmyriov, A. Shareq, K. F. Stazhilo-Alekseev, V. I. Drobov, N. A. Azimi, P. J. Gannon, B. K. Lubemov, A. Kh. Kafarskiy and E. P. Malyarov (1977) *Mineral Resources of Afghanistan*, Kabul: Ministry of Mines and Industries/ United Nations Development Programme.

Salomon, R. (1995) "On drawing socio-linguistic distinctions in Old Indo-Aryan: the question of Kṣatriya Sanskrit and related problems," in G. Erdosy (ed.), *The Indo-Aryans of Ancient South Asia: Language, Material Culture, and Ethnicity*, Berlin and New York: Walter de Gruyter.

Sāmant, R. N. and S. Kabirāj (2002) *Sāpera mantra o jhārapunka: ālocanā o samgraha* [On serpent worship and exorcism], Calcutta: Lokasamskrti Gabeshanā Parishad.

Sandfeld, K. (1930) Linguistique balkanique: problèmes et résultats. Paris, Champion.

Sankalia, H. D. (1949) *Studies in the Historical and Cultural Geography of Gujarat*, Poona: Deccan College Postgraduate and Research Institute.

—— (1974) *Prehistory and Protohistory of India and Pakistan*, Poona: Deccan College Postgraduate and Research Institute.

—— (1977) *Prehistory of India*, New Delhi: Munshiram Manoharlal.

Sankalia, H. D. and S. B. Deo (1955) *Report on the Excavations at Nasik and Jorwe, 1950–51*, Poona: Deccan College Postgraduate and Research Institute.

Sapir, E. (1949) "Male and female forms of speech in Yana," in D. Mandelbaum (ed.), *Selected Writings of Edward Sapir*, Berkeley, CA: University of California Press, pp. 206–12.

Sarkar, S. (2001) *Mother-goddess in Pre-medieval Bengal: A Study of the Evolution of Concept and Forms of Female Divinities*, Burdwan: Niharendu Aditya.

Sastri, K. A. Nilakanta – see Nilakanta Sastri, K. A.

Scherer, A. (ed.) (1968) *Die Urheimat der Indogermanen*, Darmstadt: Wissenschaftliche Buchgesellschaft.

Schoff, W. H. (1974) *The Periplus of the Erythrean Sea*, New Delhi: Oriental Books Reprint Corporation.

Schwarzberg, J. E. (1978) *A Historical Atlas of South Asia*, Chicago, IL: University of Chicago Press.

Scotton, C. M. and J. Okeju (1973) "Neighbors and lexical borrowings," *Language* **49**: 871–89.

Sebeok, T. (ed.) (1969) *Current Trends in Linguistics*, *Vol. 5: Linguistics in South Asia*, Paris: Mouton.

Sethna, K. D. (1981) *Karpasa [cotton] in Prehistoric India*, New Delhi: Biblia Impex.

Shaffer, J. G. (1984) "The Indo-Aryan invasions: cultural myth and archaeological reality, " in J. R. Lukacs (ed.), *The Peoples of South Asia*, New York and London: Plenum Press, pp. 77–90.

—— (1986) "Cultural development in the eastern Punjab," in J. Jacobson (ed.), *Studies in the Archaeology of India and Pakistan*, New Delhi: Oxford and IBH Publishing Co, pp. 195–235.

Shaffer, J. G. (1992) "The Indus Valley, Baluchistan and Helmand traditions: Neolithic through Bronze Age," in R. Ehrich (ed.), *Chronologies in Old World Archaeology*, Chicago, IL: University of Chicago Press, pp. 441–64.

Shapiro, M. C. and H. F. Schiffman (1981) *Language and Society in South Asia*, Delhi: Motilal Banarsidass.

—— (2003) "Bibliography from *Language and Society in South Asia*," Delhi: Motilal Banarsidass. Online. Available HTTP <http://ccat.sas.upenn.edu/~haroldfs/bibliogs/LSSAbiblio.htm> (8 September 2003).

Shendge, M. J. (1997) *The Language of the Harappans: From Akkadian to Sanskrit*, New Delhi: Abhinav publications.

Sherzer, J. (1976) *An Areal–Typological study of American Indian languages North of Mexico*, Amsterdam: North-Holland Pub. Co.; American Elsevier.

Sherzer, J. and R. Bauman (1972) "Areal studies and culture history: language as key to the historical study of culture contact," *SWJA* **28**: 131–52.

Shukla, S. (1981) *Bhojpuri Grammar*, Washington, DC: Georgetown University Press.

Silva-Corvalan, C. (1983) "Tense and aspect in oral Spanish narrative: content and meaning," *Language* **59**: 760–80.

Simmonds, N. W. (ed.) (1976) *Evolution of Crop Plants*, London: Longman.

Singh, J. (1974) *An Agricultural Atlas of India: A Geographical Analysis*. Kurukshetra: Vishal Publications, 1974.

Sjoberg, A.F. (1971a) "Who are the Dravidians? The present state of knowledge," in A. F. Sjoberg (ed.), *Symposium on Dravidian Civilization*, Austin & New York: Jenkins Publishing Company, The Pemberton Press, pp. 1–32.

—— (ed.) (1971b) *Symposium on Dravidian Civilization*, Austin & New York: Jenkins Publishing Company, The Pemberton Press.

Snakenberg, R. L. (1988) *The Hawaiian Sentence Book*, Honolulu, HI: The Bess Press.

Southworth, F. C. (1958) *A Test of the Comparative Method*, Ann Arbor, MI: University Microfilms.

—— (1971) "Detecting prior creolization: an analysis of the historical origins of Marathi," in D. Hymes (ed.), *Pidginization and Creolization of Languages*, Cambridge: Cambridge University Press, pp. 255–74.

—— (1974) "Linguistic stratigraphy of north India," in F. C. Southworth and M. L. Apte (eds), *Contact and Convergence in South Asian Languages, IJDL* (special issue), pp. 201–23.

—— (1975) "Sociolinguistic research in south India: achievements and prospects," in B. Stein (ed.), *Essays on South India*. University of Hawaii Press.

—— (1976a) "On subgroups in Dravidian," *IJDL* **5**: 114–37.

—— (1976b) "The verb in Marathi and Konkani," *IJDL* **5**: 298–326.

—— (1979a) "Sociolinguistic variation and semantic structure: the case of Tamil kaRi," in W. C. McCormack and S. A. Wurm (eds), *Language and Society: Anthropological Issues*, New York: Mouton Publishers, the Hague–Paris–New York, pp. 163–70.

—— (1979b) "Lexical evidence for early contacts between Indo-Aryan and Dravidian," in M. M. Deshpande and P. E. Hook (eds), *Aryan and non-Aryan in India*, Ann Arbor: The University of Michigan Center for South and Southeast Asian Studies, pp. 191–233.

—— (1980) "Functional aspects of bilingualism," *IJDL* **9**: 74–108.

—— (1988) "Ancient economic plants of South Asia: Linguistic archaeology and early agriculture," in M. A. Jazayery and W. Winter (eds), *Languages and Cultures: Studies in Honor of Edgar C. Polomé*, Berlin: Mouton De Gruyter, pp. 669–80.

—— (1990a) *Linguistic and Social Processes in Linguistic Diffusion*, in R. Filipovič and M. Bratanič (eds), *Languages in Contact: Proceedings of the Symposium 16.1. of the 12th International Congress of Anthropological and Ethnological Sciences (Zagreb July 25–27, 1988)*, Zagreb: Institute of Linguistics, Faculty of Philosophy, University of Zagreb, pp. 295–313.

—— (1990b) "The reconstruction of prehistoric South Asian language contact," in E. H. Bendix (ed.), *The Uses of Linguistics.* Annals of the New York Academy of Sciences, Vol. 583, New York: The New York Academy of Sciences, pp. 207–34.

—— (1995) "Reconstructing social context from language: Indo-Aryan and Dravidian prehistory," in G. Erdosy (ed.), *The Indo-Aryans of Ancient South Asia: Language, Material Culture and Ethnicity*, New York: Walter de Gruyter, pp. 258–77.

—— (1998) "On the origin of the word Tamiẓ," *IJDL* **26.1**: 129–32.

Southworth, F. C. and M. L. Apte (eds) (1974) *Contact and Convergence in South Asian Languages, IJDL* (special issue).

Southworth, F. C. and C. Daswani (1974) *Foundations of Linguistics*, New York: The Free Press.

Spate, O. H. K. (1954) *India and Pakistan: A General and Regional Geography*, London: Methuen & Co. Ltd.

Sresthacharya, I. (1995) *A Concise Dictionary, Newari–English*, Kathmandu: Pilgrims Book House.

Srinivasan, T. M. (1991) *Irrigation and Water Supply, South India, 200 B.C.–1600 A.D.* Madras: New Era Publications.

StII = *Studien zur Indologie und Iranistik*.

Stampe, David (2003) *Comparative Munda (mostly North), rough draft based on H-. J. Pinnow and R. D. Munda*, Online. Available HTTP: <http://ling.lll.hawaii.edu/faculty/stampe/aa.html> (30 September 2003).

Sturtevant, E. H. (1977) *The Pronunciation of Greek and Latin*, Westport, WA: Greenwood Press.

Subbarao, B. (1958) *The Personality of India*, Baroda: Faculty of Arts, M. S. University of Baroda.

Subrahmanyam, P. S. (1971) *Dravidian Verb Morphology (A Comparative Study)*, Annamalainagar: Annamalai University.

Swadesh, M. (1951) "Diffusional cumulation and archaic residue as historical explanations of language change," *SWJA* **7**.1–21.

—— (1952) "Lexico-statistic dating of prehistoric ethnic contacts," *PAPS* **96**: 453–63.

SWJA = *Southwestern Journal of Anthropology*.

TAPS = *Transactions of the American Philosophical Society*.

Thampuran, R. R. (1981) "Process of convergence and language shift: a case study of the Kudumbis of Kerala," in *Seminar on Convergence, Pidginization & Simplification of Languages*, Mysore: Central Institute of Indian Languages.

Thapar, R. (1966) *A History of India-I*, Hammondsworth: Penguin Books.

—— (1973) "Image of the barbarian in early India," *Comparative Studies in Society and History* **13**: 408–36. (Republished in R. Thapar (1978) *Ancient Indian Social History: Some Interpretations*, New Delhi: Orient Longman.)

—— (1975) "Puranic lineages and archaeological cultures," *Puratattva* **8**: 86–98. (Republished in R. Thapar (1978) *Ancient Indian Social History: Some Interpretations*, New Delhi: Orient Longman.)

—— (1978) *Ancient Indian social history: Some interpretations*, New Delhi: Orient Longman.

Thapar, R. (1984) *From Lineage to State*, Bombay: Oxford University Press.
—— (2002) *Early India: From the Origins to AD 1300*, London: Allen Lane, The Penguin Press.
Thieme, P. (1955) "Review of *The Sanskrit Language*, by T. Burrow," *Language* **31**: 428–48.
Thomas, P. K. and P. P. Joglekar (1994) "Holocene faunal studies in India," *Man and Environment* **19**: 179–203.
Thurston, E. (1909) *Castes and Tribes of Southern India*, Madras: Government Press.
Tikkanen, B. (1988) "On Burushaski and other ancient substrata in northwest South Asia," *Studia Orientalia* **64**: 303–25.
Tiwari, U. N. (1960) *The Origin and Development of Bhojpuri*, Calcutta: The Asiatic Society.
TPS = *Transactions of the Philological Society.*
Trager, G. L. (1939) "'Cottonwood' = 'tree': a southwestern linguistic trait," *IJAL* **9**: 117–18. (Reprinted in D. Hymes (ed.) (1964) *Language in Culture and Society: A Reader in Linguistics and Anthropology*, New York: Harper & Row Publishers, pp. 467–8.
Trautmann, T. R. (1981) *Dravidian Kinship*, Cambridge: Cambridge University Press.
Troubetzkoy, N. S. (1939) "Gedanken über das Indogermanenproblem," *Acta Linguistica* **1**: 81–9.
Trumpp, E. (1970) *Grammar of the Sindhi Language*, Osnabrück: Biblio Verlag (first published 1872).
Tulpule, S. G. (1960) *An Old Marathi Reader*, Poona: Venus Prakashan.
Turner, Sir R. L. (1966) *A Comparative Dictionary of the Indo-Aryan Languages*, London: Oxford University Press.
—— (1975) *Collected Papers (1912–1973)*, London: Oxford University Press.
Tyler, S. A. (1968) "Dravidian and Uralian: The lexical evidence," *Language* **44**: 798–812.
United States Office of Geography (1952) *India: Official Standard Names Gazetteer*, Vols. I–II, Washington, DC: Division of Geography, Department of the Interior.
Unni, K. R. (1959) Caste in southern Malabar, unpublished PhD dissertation, M. S. University of Baroda.
Vats, M. S. (1940) *Excavations at Harappa*, Delhi: Government of India.
Vavilov, N. I. (1951) The origin, variation, immunity and breeding of cultivated plants: selected writings of N. I. Vavilov; translated from the Russian by K. Starr Chester. Waltham (MA): Chronica Botanica Co.
Vishnu-Mittre (1969) "Remains of rice and millet," in H. D. Sankalia, S. B. Deo and Z. D. Ansari, *Excavations at Ahar (Timbavati)*, Poona: Deccan College Postgraduate and Research Institute.
—— (1974) "The beginnings of agriculture: palaeobotanical evidence in India," in Sir J. Hutchinson (ed.), *Evolutionary Studies in World Crops: Diversity and Change in the Indian Subcontinent*, Cambridge: Cambridge University Press, pp. 3–30.
Vishnu-Mittre and R. Savithri (1982) "Food economy of the Harappans," in G. Possehl, (ed.), *The Harappan Civilization: A Contemporary Perspective*, New Delhi: Oxford & IBH Publishing Co., pp. 205–22.
Vogel, J. P. (1926) *Indian Serpent-lore, or, the Nāgas in Hindu Legend and* Art, London: A. Probsthain.
Warmington, E. H. (1928) *The Commerce between the Roman Empire and India*, Cambridge: The Cambridge University Press.
Weber, S. (1991) *Plants and Harappan subsistence: An example of stability and change from Rojdi*, Boulder: Westview Press.

Weinreich, U. (1974) *Languages in contact: Findings and problems*, Paris: Mouton.

Wheeler, S. M (1968) *The Indus Civilization*, Cambridge: Cambridge University Press.

Whitney, W. D. (1950) *Sanskrit Grammar*, Cambridge, MA: Harvard University Press, Oxford: Oxford University Press.

Winternitz, M. M. (1959) *A History of Indian Literature*, Calcutta: University of Calcutta.

Witzel, M. (1989) "Tracing the Vedic dialects," in C. Caillat (ed.), *Dialects dans les littératures Indo-Aryennes*, Paris: Institut de Civilisation Indienne, pp. 97–264.

—— (1993) "Nepalese hydronomy: towards a history of settlement in the Himalayas," in G. Toffin (ed.), *Nepal, Past and Present: Proceedings of the Franco-German Conference, Arc-et-Senans, June 1990*, New Delhi: Sterling Publishers.

—— (1995a) "Early Indian history: linguistic and textural parameters," in G. Erdosy (ed.), *The Indo-Aryans of Ancient South Asia: Language, Material Culture and Ethnicity*, Berlin: Walter de Gruyter, pp. 85–125.

—— (1995b) "Rgvedic history: poets, chieftains and polities," in G. Erdosy (ed.), *The Indo-Aryans of Ancient South Asia: Language, Material Culture and Ethnicity*, Berlin: Walter de Gruyter, pp. 307–352.

—— (1995c) "Early sanskritization: origins and development of the Kuru state," *EJVS* **1–4**: 1–26.

—— (1999a) "Aryan and non-Aryan Names in Vedic India. Data for the linguistic situation, c. 1900–500 B.C.," in J. Bronkhorst and M. M. Deshpande (eds), *Aryan and Non-Aryan in South Asia: Evidence, Interpretation and Ideology*, Cambridge (MA): Harvard University Press, pp. 337–404.

—— (1999b) "Substrate languages in old Indo-Aryan," *EJVS* **5**: 1–67. (References herein use the page numbers of the .pdf version to be found at <http://www1.shore.net/~india/ejvs/ejvs501article.pdf> (15 August 2003).

—— (2001) "Autochthonous Aryans? The evidence from Old Indian and Iranian texts," *EJVS* **7–3**: 1–93.

Zide, A. K. and N. H. Zide (1973) "Semantic reconstructions in proto-Munda cultural vocabulary I," *Indian Linguistics* **34**: 1–24.

Zide, N. H. and A. K. Zide (1976) "Proto-Munda cultural vocabulary: evidence for early agriculture," in P. Jenner, L. Thompson and S. Starosta (eds), *Austro-Asiatic Studies: Papers from the First International Conference on Austro-Asiatic Linguistics, Honolulu, 1973*, Honolulu, HI: University Press of Hawaii.

Zide, N. H. (1969) "Munda and non-Munda Austro-Asiatic Languages," in Thomas Sebeok (ed.), *Current Trends in Linguistics*, The Hague: Mouton, pp. 411–30.

Zvelebil, K. (1970) *Comparative Dravidian Phonology*, The Hague: Mouton.

—— (1977) *A Sketch of Comparative Dravidian Morphology: Part I*, The Hague: Mouton Publishers.

—— (1990) *Dravidian linguistics: An Introduction*, Pondicherry: Pondicherry Institute of Linguistics and Culture.

INDEX

The following is a comprehensive index, including all proper names (of authors and scholars, places, literary works), languages and language families, key words for topics discussed in the preceding pages, and the English meanings of reconstructed Dravidian words (Chapter 8 and Appendices A–C) and other relevant words in South Asian languages. Languages are identified by the abbreviation "lg." following the language name, or by the name of the language family or subfamily in brackets. ("Dravidian" is abbreviated as "Dr."; *see* pp. xi–xiv for other abbreviations.) Names of flora are indexed under the English or Anglo-Indian name.

Page numbers in **bold** type refer to definitions or key discussions of topics. Underlining indicates words in South Asian languages: thus the entry "chameleon 258, 272" refers only to words in South Asian languages which mean "chameleon", while the entry "adze 246, 266" indicates that a word meaning "adze" is to be found on page 266. Most such entries refer to Dravidian (PD, PSD, or PSD1) reconstructions to be found on pages 257–83.

Readers may also note that each chapter begins with a detailed table of contents.